RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Continuous evolution of user-defined genes at 1 million times the genomic mutation rate

Gordon Rix, Rory L. Williams, Vincent J. Hu, Aviv Spinner, Alexander (Olek) Pisera, Debora S. Marks, Chang C. Liu*

INTRODUCTION: When a gene evolves under prevailing and shifting functional demands, those demands become embedded into the resulting diversity of homologous sequences as patterns of conservation and change. We have long used these patterns to extract features important for gene function and to infer how genes historically changed to satisfy new demands, offering lessons in biomolecular design. Yet only natural evolution, through millions of years of operation, has consistently diverged genes far enough to yield sets of homologous sequences containing abundant statistical information on the details of their function. Compressing long-term gene evolution to laboratory time spans at scale would break nature's monopoly on the production of extensive gene diversity and would enable the systematic detection of additional structural and functional demands governing biology, the engineering of custom biomolecules, and the prospective study of evolutionary mechanisms and principles by which gene diversity was generated in the first place.

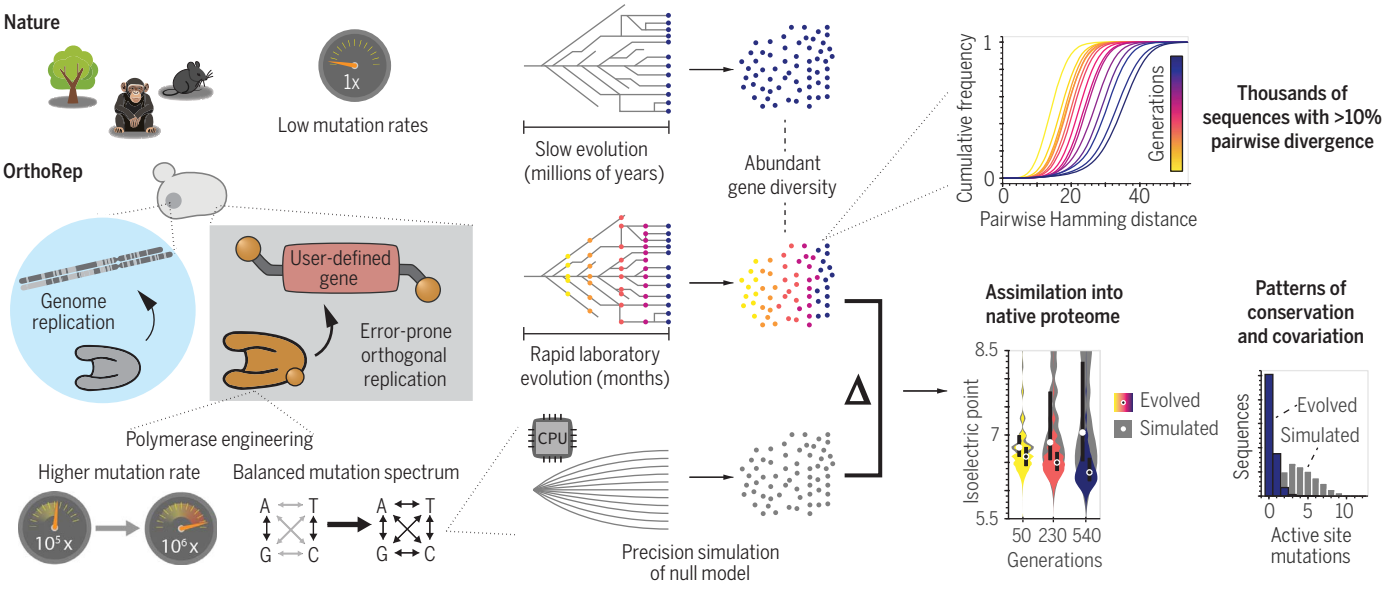
RATIONALE: Orthogonal DNA replication (OrthoRep) is a genetic architecture for the continuous hypermutation of user-defined genes in vivo, but the mutation rates of legacy systems are too low to condense long gene evolutionary trajectories onto laboratory timescales when strong directional selection is absent. OrthoRep systems that intensify the mutational force on chosen genes should enable their extensive laboratory evolution toward nature-like levels of diversity under all types of selection.

RESULTS: We engineered OrthoRep systems with mutation rates exceeding 10^{-4} substitutions per base. We encoded a maladapted tryptophan synthase subunit, TrpB, onto OrthoRep in *Saccharomyces cerevisiae* and continuously evolved it to gain and maintain the ability to synthesize tryptophan over ~540 generations in 96 independent populations. TrpB diverged extensively. The median distance separating pairs of evolved sequences reached 35 amino acids (~9%) with thousands of distinct pairs separated by >60 amino acids (15%). For comparison, the me… distance between mouse and human ort… ogous genes is ~11%. The rich collection of diverged TrpB sequences revealed known and unexpected factors influencing TrpB's function and evolution. Varying degrees of conservation consistent with structural principles, regions and networks of mutations responsible for functional adaptation, signatures of thermoadaptation, and a trend toward net negative charge hypothesized to avoid indiscriminate macromolecular clustering inside cells were inferred by comparing evolved TrpBs with a precise null model of sequence change. The high fitnesses of extensively diverged TrpB variants were not predictable by a state-of-the-art machine learning model trained on natural variation.

CONCLUSION: OrthoRep's ability to condense long adaptive and neutral gene evolutionary processes into accessible laboratory experiments should drive the evolutionary engineering of new biomolecules, the broader mapping of fitness landscapes, and the extraction of biological forces governing the functions of genes and biomolecules in vivo. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ccl@uci.edu
Cite this article as G. Rix *et al.*, *Science* **386**, eadm9073 (2024). DOI: 10.1126/science.adm9073

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adm9073



OrthoRep compresses extensive gene evolution onto laboratory timescales and supports the detection of structural and functional demands acting on chosen genes. (Left) Gene sequence divergence occurs slowly in nature but rapidly through continuous evolution with high–mutation rate OrthoRep systems. (Center) Abundant gene diversity evolved with OrthoRep can be used to detect constraints shaping a gene's structure and function through comparison against a null model of sequence change based on OrthoRep's fully characterized mutation rate and preferences. (Right) Extensive sequence divergence under selection was experimentally achieved for a model gene, trpB. Patterns of sequence change over time enabled the inference of essential positions, functional adaption in a subdomain of TrpB, charge and temperature optimization, and networks of functionally interconnected residues. CPU, central processing unit.

RESEARCH ARTICLE

SYNTHETIC BIOLOGY

Continuous evolution of user-defined genes at 1 million times the genomic mutation rate

Gordon Rix[1], Rory L. Williams[2], Vincent J. Hu[2], Aviv Spinner[3], Alexander (Olek) Pisera[2], Debora S. Marks[3,4], Chang C. Liu[1,2,5,6]*

When nature evolves a gene over eons at scale, it produces a diversity of homologous sequences with patterns of conservation and change that contain rich structural, functional, and historical information about the gene. However, natural gene diversity accumulates slowly and likely excludes large regions of functional sequence space, limiting the information that is encoded and extractable. We introduce upgraded orthogonal DNA replication (OrthoRep) systems that radically accelerate the evolution of chosen genes under selection in yeast. When applied to a maladapted biosynthetic enzyme, we obtained collections of extensively diverged sequences with patterns that revealed structural and environmental constraints shaping the enzyme's activity. Our upgraded OrthoRep systems should support the discovery of factors influencing gene evolution, uncover previously unknown regions of fitness landscapes, and find broad applications in biomolecular engineering.

Over the history of life, evolution has carried out a large-scale experiment exploring how gene sequences change under the constraints of prevailing or shifting structural and functional demands. The results of this natural experiment, embedded within the patterns of diversity across extant gene sequences, are of fundamental value to almost all areas of life sciences. For example, sequence conservation within a gene family is used to identify functionally critical residues (*1*–*4*), covariation among positions in an RNA or protein is used to deduce structural contacts and sectors of connectivity (*5*–*12*), differences in amino acid composition reveal environmental preferences [e.g., temperature (*13*–*15*) or subcellular localization (*16*)], and differences in the conserved physicochemical properties across regions of a protein reflect the driving forces behind folding (*17*, *18*). Natural diversity across homologs also serves as a shared biomolecular engineering resource that can be mined for desired activities or recombined to access new functions (*19*–*25*). Additionally, machine learning (ML) models have proven incredibly effective at extracting meaningful representations of biomolecular structure and function from the extensive diversity within and across gene families, as exemplified by the ability of ML models to predict functional effects of mutations (*26*–*28*), design functional sequences (*29*–*31*), and predict protein structures (*32*–*34*). However, the natural evolution of highly diverse gene sequences under the constraints of selective forces—or conversely, the imprinting of selective forces and design principles into the statistics of sequence diversity—takes a long time at the slow rates of mutation in cellular and multicellular organisms. For example, reaching the 11% median divergence separating essential mouse and human genes (*35*) took ~96 million years (*36*). Moreover, generating extensive collections of diverged sequences required complex histories of geographical isolation and speciation to maintain variation across populations in the face of within-population selective sweeps or genetic drift. Is it possible to compress long and vast gene evolution processes into laboratory experiments? Doing so would allow us to systematically detect previously unknown structural and functional constraints governing biology, engineer custom biomolecules, create rich sources of genetic variation for neofunctionalization or ML, and prospectively study the mechanisms and principles by which histories of selective forces become embedded into the patterns of sequence diversity.

We and others have endeavored toward this goal (*37*, *38*), including through the development of scalable accelerated continuous evolution systems (*39*, *40*), such as our orthogonal DNA replication (OrthoRep) system in *Saccharomyces cerevisiae* (*41*, *42*). OrthoRep cells have an additional DNA replication system comprising an orthogonal DNA polymerase (DNAP)–plasmid pair wherein the orthogonal DNAP (TP-DNAP1) durably replicates the orthogonal plasmid (p1) but not the host genome; likewise, host DNAPs replicate the host genome but not p1 (Fig. 1A). Through this architecture, OrthoRep supports the sustainable coexistence of two independent mutation rates in the same cell: a low mutation rate of 10^{-10} substitutions per base (s.p.b.) for the host genome and a high mutation rate of 10^{-5} s.p.b. for p1. OrthoRep's 10^{-5}-s.p.b. mutation rate exceeds the error thresholds of the large host genome but not those of the small p1 plasmid (*42*), which allows us to drive the rapid, continuous evolution of p1-encoded chosen genes as cells autonomously propagate. However, 10^{-5} s.p.b. is still not high enough to observe extensive evolution on laboratory timescales in the general case where evolution occurs both with and without positive selection. In the specific case of evolution under strong positive selection, sufficiently large population sizes can directly compensate for moderate mutation rates by increasing the beneficial mutation supply on which selection "pulls"; in this case, OrthoRep systems have successfully evolved enzymes (*38*, *43*, *44*), biosynthetic pathways (*45*), biosensors (*46*), drug targets (*42*), and antibodies (*47*, *48*) through long adaptive mutational pathways. Yet, in the general case that includes when purifying selection is dominant or when selection is absent—both highly relevant in the generation of natural diversity and the ability to escape local fitness optima—mutation becomes the main force pushing sequence change. Without the pull of positive selection, 10^{-5}-s.p.b. OrthoRep systems would take 100 generations (8 to 12 days for the yeast host of OrthoRep) just to observe an average of one new mutation in a typical 1-kb gene.

In this work, we present OrthoRep systems with a mutation rate up to 1.7×10^{-4} s.p.b., corresponding to a new mutation in a typical 1-kb gene once every <10 generations in the absence of any selection. This intensified mutational force on chosen genes, operating at >1 million times the mutation rate of the host genome, allows us to mimic extended periods of natural gene evolution on laboratory timescales in the general case. We show how, in ~3 months of laboratory passaging (totaling <15 hours of researcher intervention) of 96 independent populations, a conditionally essential gene encoded on p1 diverges to an extent where the median distance separating pairs of evolved sequences is 35 amino acids, with thousands of distinct pairs separated by >60 amino acids. This corresponds to an amino acid divergence of ~9% to >15%, exceeding the median 11% distance between mouse and human orthologous genes (*35*). By analyzing the rich collection of diverged sequences throughout their laboratory evolutionary history, we uncover hidden forces shaping and constraining sequence change, such as a preference for negative net charge, supporting a proposed mechanism by which proteins avoid large-scale indiscriminate clustering in the crowded environment inside cells (*49*). We also extract examples of allosteric network remodeling through the

[1]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92617, USA. [2]Department of Biomedical Engineering, University of California, Irvine, CA 92617, USA. [3]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [4]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. [5]Department of Chemistry, University of California, Irvine, CA 92617, USA. [6]Center for Synthetic Biology, University of California, Irvine, CA 92617, USA.
*Corresponding author. Email: ccl@uci.edu

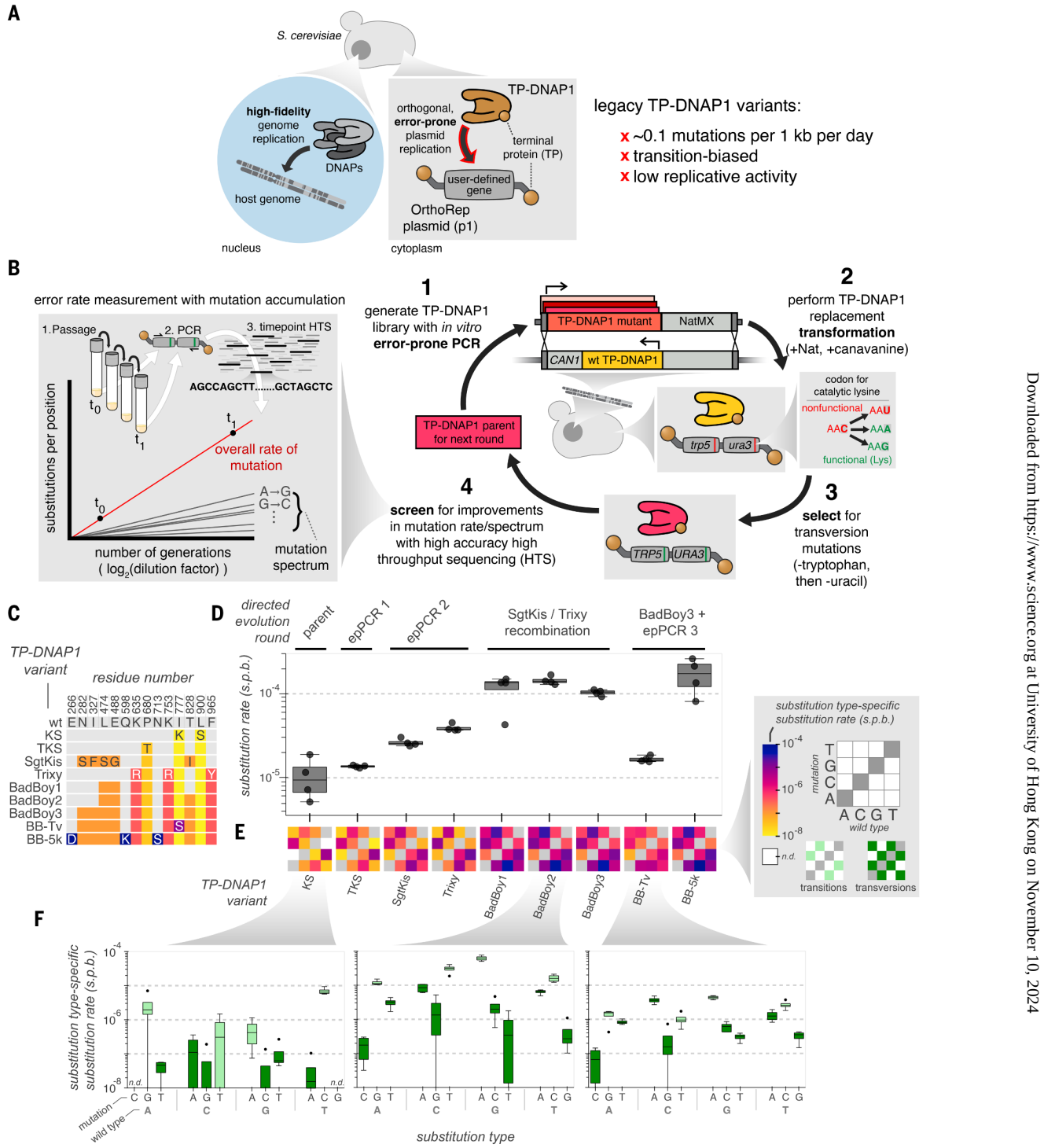


Fig. 1. Engineering orthogonal DNA polymerases for increased mutation rates. (**A**) Architecture of the OrthoRep system. A DNA polymerase (TP-DNAP1) that exclusively replicates a specific cytoplasmically localized plasmid through protein-primed replication at a high error rate enables in vivo targeted mutagenesis without mutagenizing genomic DNA. (**B**) Schematic for a directed evolution approach to engineer TP-DNAP1's mutation rates and mutation spectrum incorporating both a direct selection for rare transversion mutations as well as high-accuracy mutation rate measurement using a mutation accumulation and HTS assay. (**C**) Mutations identified in TP-DNAP1 variants presented in this study. Color and single-letter amino acid code are used for only the first TP-DNAP1 in which a mutation is identified. Color only is shown for all other instances of a mutation. (**D** to **F**) Mutation rate measurements via mutation accumulation for a series of TP-DNAP1 directed evolution intermediates showing either overall mutation rates as boxplots (D), mean mutation rates for individual substitution types as heatmaps (E), or mutation rates for individual substitution types for three individual TP-DNAP1 variants as boxplots (F). Points are representative of individual biological replicates, each representing three to four time points with >50 sequences each. Boxplots and heatmaps are representative of $n = 4$ biological replicates. Box plot central line, boxes, and whiskers represent the median, interquartile range, and minimum and maximum values, respectively. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

cooccurrence of mutations across distinct clades and temperature optimization through amino acid content change. Overall, our work provides an approach to systematically reveal the evolutionary constraints and selective forces that genes experience and delivers an upgraded OrthoRep system for broad application.

Results

OrthoRep engineering

The current state-of-the-art OrthoRep system uses TP-DNAP1-4-2 as the error-prone orthogonal DNAP (*42*). Besides its suboptimal error rate of 10^{-5} s.p.b., TP-DNAP1-4-2 also has low replicative activity (fig. S1) and exhibits a heavily transition-biased mutation spectrum (fig. S2) (*42*), suppressing the impact of point mutations on amino acid sequence during protein evolution (fig. S3). We carried out a directed evolution campaign on TP-DNAP1 to increase OrthoRep's overall error rate, transversion rate, and activity. A selection strain, OR-Y488, was engineered to contain a p1 plasmid (p1-ura3*-trp5*) encoding two auxotrophic marker genes, *ura3* and *trp5*, each specifically disabled through an active site missense mutation whose sole option for functional reversion is a transversion (Fig. 1B and fig. S4). TP-DNAP1s with the highest transversion rates should restore *URA3* and *TRP5* most frequently, resulting in their enrichment from genomically integrated TP-DNAP1 libraries (Materials and methods and fig. S5) when OR-Y488 is grown in the absence of exogenous uracil or tryptophan. Selection was designed to occur in two sequential stages, first for *URA3* restoration and then for *TRP5* restoration, to suppress the enrichment of low–error rate TP-DNAP1 variants in revertants that stochastically emerge from the long tail of the Luria-Delbrück distribution (*50*).

To precisely guide the TP-DNAP1 directed evolution campaign, we developed a mutation accumulation assay (*51*) for p1 and coupled it with high-throughput sequencing (HTS) of p1 amplicons using the Oxford Nanopore Technologies (ONT) platform (*52*), allowing us to accurately determine the rate for any individual type of mutation (Fig. 1B). In this assay, a strain containing p1 replicated by a given TP-DNAP1 variant is grown for a set number of generations. A region of p1 not under selection is sequenced at two or more time points using unique molecular identifiers (UMIs) for error correction (*53*, *54*), and the rate of change in the number of mutations per position is calculated individually for all types of mutations to fully describe the overall mutation rate and mutation preferences of the TP-DNAP1 variant (Materials and methods). To facilitate rapid characterization, we developed a custom analysis pipeline that could carry out most of the analysis steps autonomously (fig. S6). This pipeline, Mutation Analysis for Parallel Laboratory Evolution (or Maple), performs consensus sequence generation, demultiplexing, mutation identification, mutation rate analysis, and many other operations to generate a collection of visualizations and data tables that accelerate analysis of mutation-rich sequencing datasets while minimizing user input.

With the genetic selection, HTS-based mutation accumulation measurement pipeline, and Maple in place, we carried out the TP-DNAP1 directed evolution campaign over five rounds (Fig. 1B), yielding a collection of TP-DNAP1 variants (Fig. 1C and table S1) with broad mutational spectra and error rates up to and exceeding 10^{-4} s.p.b. (Fig. 1, D to F). The full course of the directed evolution campaigns is described in the supplementary text and figs. S7 to S13. We emphasize three key outcomes. First, because our mutation rate measurement pipeline obtained complete mutation rate and preference data for our evolved TP-DNAP1 variants, the use of OrthoRep in driving continuous gene evolution experiments comes with the ability to generate accurate null models of sequence change in the absence of selective forces against which evolutionary trajectories and outcomes can be compared. Second, we saw evidence suggestive of nearing gene error thresholds at OrthoRep's 10^{-4}-s.p.b. mutation rates. We found that mutation frequencies were consistently higher in regions of p1 not under selection compared with regions encoding genes under selection (fig. S14, A and B) and that the ratio between the two was highest for the TP-DNAP1 variants with the highest mutation rates (fig. S14C). This implies that the 10^{-4}-s.p.b. mutation rates of TP-DNAP1 variants BadBoy1, BadBoy2, and BadBoy3 (Fig. 1D) quickly degrade the function of genes when purifying selection is absent. Furthermore, an extremely error-prone 1.7×10^{-4}–s.p.b. TP-DNAP1 variant that we isolated (BB-5k) did not durably maintain p1 in two of four biological replicates under selection over ~120 generations of mutation accumulation, possibly because BB-5k exerted an excessive mutational load on the selection marker used to maintain p1, leading to mutational meltdown. Third, in conjunction with past efforts, the TP-DNAP1s we obtained complete a set of OrthoRep systems evenly spanning a range of ~5 orders of magnitude, from ~10^{-9} s.p.b.—similar to the mutation rate of modern cellular genomes—up to ~10^{-4} s.p.b., which is likely in the regime where the error thresholds of individual genes reside, above which gene mutational meltdown occurs and near which maximal gene adaptation rates can be reached (*55*, *56*). Thus, the overall ecosystem of OrthoRep systems should allow for not only the rapid continuous evolution of chosen genes in vivo but also investigations on the detailed role of mutation rates and error thresholds in molecular evolution, for which theory is abundant but experiment is sparse.

Extensive divergence of a conditionally essential gene on laboratory timescales

Our OrthoRep systems should be capable of driving rapid evolution of chosen genes regardless of the type of selection imposed. We encoded the β subunit of *Thermotoga maritima*'s tryptophan synthase (TrpB) onto p1 in a tryptophan auxotroph where p1 is exclusively replicated by BadBoy2 at a mutation rate of 1.4×10^{-4} s.p.b. (Fig. 2A) (*57*–*59*). TrpB condenses indole with serine to yield tryptophan (Trp), but *T. maritima* TrpB is maladapted for this standalone reaction because it normally functions in complex with TrpA (*57*, *58*). Therefore, cells grown in the absence of Trp need to evolve improved TrpB activity to propagate, which allows TrpB to serve as the subject of an extended evolution experiment that includes a range of selection pressures (Fig. 2B and table S2).

We designed our evolution experiment to prioritize sequence divergence and diversity to maximize the amount of evolutionary information that could later be extracted. The evolution experiment was therefore run for ~540 generations (~3 months) at a scale of 96 independent replicate 500 μl cultures. In total, 1:1024 (10 generation) transfers into fresh growth medium were made every 1 to 3 days, depending on cell density, following a passaging schedule that included all types of selection pressures (Fig. 2B): an initial period of evolution without selection ("no selection"); then a period of adaptation when strong selection pressure was applied and functional improvements in TrpB's function were observed ("mostly positive selection"); followed by a long period characterized by mostly purifying selection, where adapted TrpBs were pressured to maintain the fitness that they evolved ("mostly purifying selection") (Fig. 2B). The mostly purifying period also included some brief episodes of relaxed or removed selection pressure, which we introduced with the intention of promoting sequence divergence. We collected DNA from cells at 15 time points throughout the ~540-generation evolution experiment (Fig. 2B) and used a rolling circle amplification–based sequencing strategy in conjunction with HTS and Maple to analyze the TrpB sequences sampled from these time points (Materials and methods and fig. S15).

Overall, we observed a monotonic increase in both the average number of mutations and diversity (as measured by pairwise Hamming distances) in TrpB throughout all phases of evolution (Fig. 2, C to E), resulting in a large number of distinct evolutionary outcomes (fig. S16). The rate at which mutations accumulated in the population was highest in the no selection period (~0.15 amino acid changes per generation) followed by the mostly positive selection and mostly purifying selection periods (~0.039 and ~0.021 amino acid changes per generation, respectively) (table S3). The

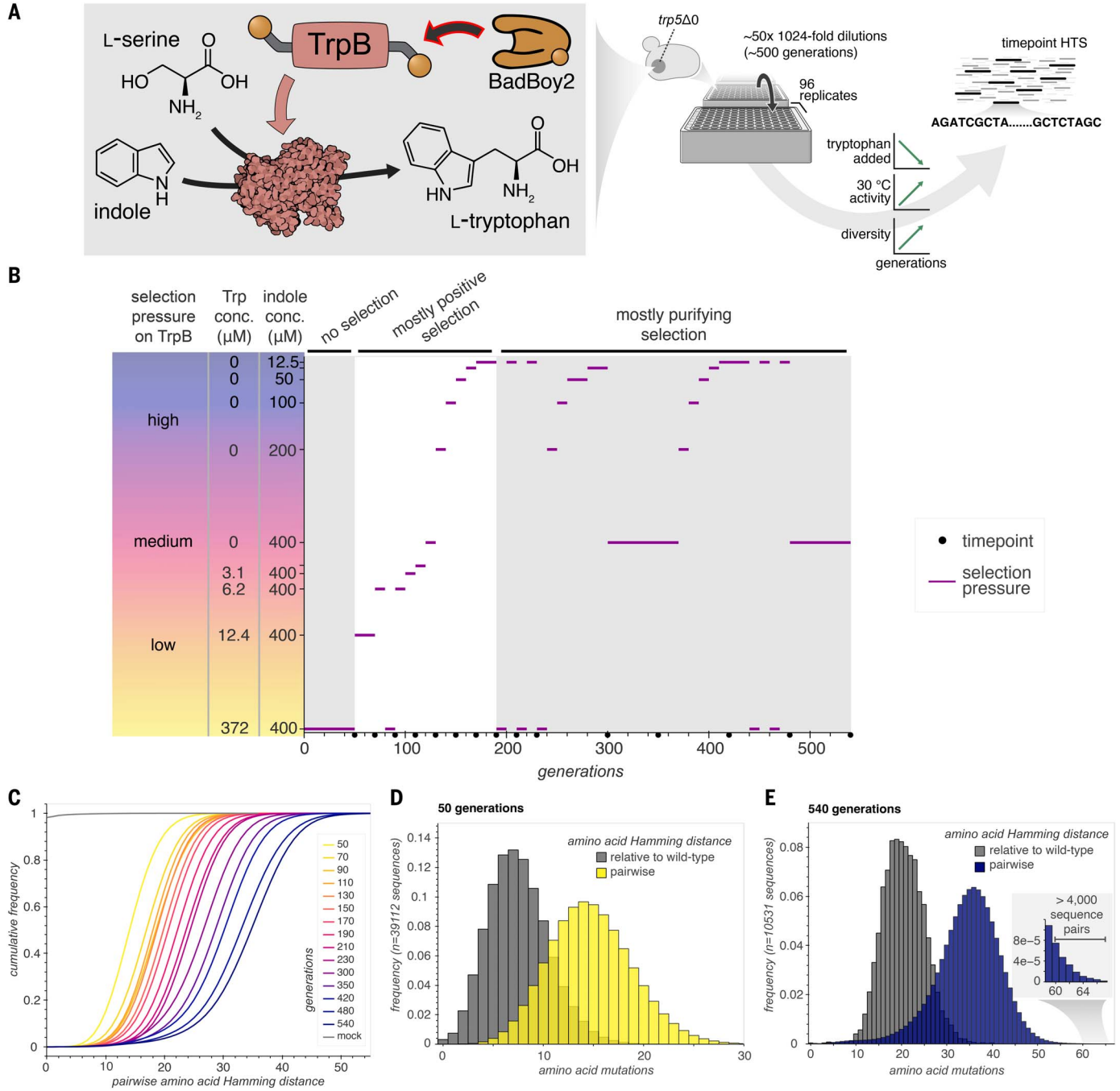


Fig. 2. Massively parallel continuous diversification and evolution of TrpB. (**A**) Schematic for OrthoRep-driven evolution of the tryptophan synthase β subunit from *T. maritima* (TrpB) for standalone function in yeast. TrpB was integrated onto the p1 plasmid in a yeast strain lacking the native yeast tryptophan synthase gene (*TRP5*). Ninety-six independent cultures of the resulting strain were passaged mostly under selective pressure for Trp production using exogenously supplied indole over ~540 generations. DNA from 15 time points throughout the evolution campaign was harvested and sequenced using HTS. TrpB illustration generated using Illustrate (*59*). (**B**) Selection pressure for TrpB function is applied by lowering or eliminating exogenously supplied Trp and lowering exogenously supplied indole over time. The schedule of selection pressure imposed throughout extensive evolution is plotted. Time points at which cultures were harvested and sequenced using HTS are indicated. Selection periods were characterized as no selection, mostly positive selection, and mostly purifying selection based on the Trp and indole amounts supplied and the progress of evolution. See Materials and methods and table S2 for a description of media conditions and selection pressure derivation. (**C** to **E**) Distribution of amino acid Hamming distances for both pairwise comparisons and comparisons with the WT sequence, at the first and last time point [(D) and (E), respectively] and as pairwise cumulative distributions for all time points (C).

mostly positive selection period did not have the highest rate of mutation accumulation even though there was substantial adaptation in producing TrpBs that supported cell growth in the absence of Trp and presence of moderate concentrations of indole. This suggests that BadBoy2's error rate was high enough to consistently "saturate" positive selection with an overabundance of beneficial mutations in TrpB, predicting the broader power of upgraded OrthoRep mutation rates in biomolecular evolution applications. Mutation accumulation in

the mostly purifying selection period was appreciable yet substantially slower than in the no selection and mostly positive selection periods. This suggests that BadBoy2's error rate was high enough to constantly test the constraints of structure and function, signaling the general potential of OrthoRep in uncovering biological forces governing how genes and biomolecules operate. At the end of the evolution experiment, each TrpB sequence had an average of 20.6 amino acid and 44.5 nucleotide mutations (table S3). In the last two time points, >1800 sequences (~5%) had accumulated >30 amino acid changes from the ancestral 398–amino acid wild-type (WT) TrpB. The distribution of pairwise Hamming distances showed that sequences had substantially diverged from each other (Fig. 2, C and D) such that in the final time point, 24% of sequences were separated from each other by 40 amino acids or more, including more than 4000 sequence pairs differing by a pairwise amino acid Hamming distance of at least 60 (Fig. 2E, inset). This level of sequence divergence (>15%) approximates that between human and mouse essential gene orthologs (*35*), demonstrating that we can extensively evolve a gene over long mutational pathways shaped by varied selection conditions on laboratory timescales. Can the diversity of sequences be analyzed to extract structural, functional, environmental, and evolutionary information about the gene?

General structural and functional constraints

Approximately 500,000 sequences of TrpB with an average of 13.1 amino acid replacements each were captured over the evolution experiment, and >90% of those sequences were distinct (table S3). With such a diverse evolutionary dataset, patterns of conservation should contain structural and functional constraints defining TrpB. To test this notion, we used an AlphaFold structure (*33*) and knowledge from previous studies on TrpB (*60*, *61*) to first categorize each residue in TrpB according to its general structural or functional role, as outlined in Fig. 3A. We then examined whether different categories showed different levels of conservation. We immediately noticed a congruence between relative conservation and buried residues, revealing the well-known importance of a buried hydrophobic core in protein folding (Fig. 3B). We also noticed that residues within 5 Å of TrpB's active site were highly conserved. Additionally, there was a relative abundance of amino acid replacements at certain positions in the COMM domain, suggesting that it was a target of adaptation.

To improve our resolution in such observations, we generated a simulated dataset of TrpB mutants where each sequence accumulated mutations from the exact encoding of WT TrpB using BadBoy2's fully described mutation biases. For each simulated sequence, mutation accumulation was stopped once it matched the number of synonymous mutations of a corresponding sequence from the real dataset. The simulated dataset serves as the null model, where patterns in evolved TrpB sequences are simply a reflection of the mutation preferences of BadBoy2 and codon usage of WT TrpB. Under the assumption that synonymous mutations have no effect on fitness—we confirm this assumption for the experiment at hand in fig. S17—the nonsynonymous differences between the real dataset and the simulated dataset contain the influence of selective forces. An excess of nonsynonymous changes in the real dataset compared with the simulated dataset is therefore an indication of positive selection, whereas the opposite signifies purifying selection. As shown in Fig. 3C, at the generation 70 time point, the real dataset has a paucity of nonsynonymous mutations per sequence in the active site region and buried residues and an excess of nonsynonymous mutations per sequence in the COMM domain. Generation 70 is during the first phase of positive selection for TrpB's operation as a standalone enzyme capable of generating tryptophan (Fig. 2B), so this time point is most likely to reveal signatures of adaptation. (This is supported by fig. S18, which shows the overall mutation accumulation dynamics; generation 70 is the time point with the greatest excess of nonsynonymous mutations relative to the simulated dataset among the time points for which selection for TrpB function had been applied.) At generation 70, we find that positive selection had enriched mutations to buried and COMM domain residues while purifying selection had already removed changes to the active site region. Our detection of the COMM domain as a focus of adaptation can be explained because TrpB is a well-studied enzyme. The COMM domain mediates allosteric activation of TrpB by TrpA (*57*). Because our evolution experiment required TrpB to operate in a standalone manner without TrpA, the remodeling of allosteric networks through the COMM domain was an expected means to adaptation in line with previous studies of engineered TrpB standalone activity (*60*, *61*). Had TrpB not been well studied, the detection of this critical region for adaptation from the evolutionary information could have suggested such an explanation.

We also considered the final time point of the evolution experiment in detail. By generation 540, the influence of purifying selection had dominated, as evidenced by the paucity of mutations in the real data compared with the simulated (Fig. 3C) or, similarly, the consistently low rate of nonsynonymous versus synonymous mutation accumulation after generation 90 (figs. S17 and S18). Although purifying selection constrained all regions of TrpB, some regions were clearly more constrained than others. For example, the active site region had almost no mutations (maximally 1 or 2 but mostly 0) and deviated from the simulated mutant distribution more than all other regions. Buried residues also had substantially fewer nonsynonymous mutations than the simulated dataset. By contrast, the effect of purifying selection was less pronounced on surface residues, reflecting the relative tolerance of protein surfaces to mutation. This also applied to the newly solvent-exposed β-α interface region. In the absence of the α subunit, this region should be more solvent exposed than in TrpB's native context. The fact that there was little noticeable difference in the effects of selection on the β-α and β-β interfaces suggests that solvent exposure of this region had minimal impact on TrpB fitness.

Isoelectric point evolution for intracellular compatibility

We examined the 20 residues that were mutated in greatest excess across real evolved sequences relative to simulated sequences in the first two time points (Fig. 3D). Despite the entire WT TrpB protein containing only 19 arginine residues in total (5%), arginine constituted 8 of these 20 most frequently mutated residues by generation 50. This enrichment occurred even before selection for TrpB function was imposed (Fig. 2B). This led us to hypothesize charge optimization as a driving selective force because charge could influence not only TrpB function itself but also the cellular environment within which TrpB operated. To evaluate this hypothesis, we calculated the isoelectric point (pI) of sequences throughout the evolution experiment and examined its distribution over time (Fig. 3E). We found that the pI of sequences was significantly lower at the end of the experiment compared with early in the experiment ($P < 0.0001$, Mann-Whitney U test). Comparison of pI change against simulation corroborates that this effect was driven by positive selection. A similar analysis of hydrophobicity revealed a modest decrease in hydrophobicity over time for simulated sequences that was mitigated by selection in the real data (fig. S19). The change in hydrophobicity for the real sequences throughout the experiment was less pronounced than the change in pI, however, which suggests that charge optimization—and not polarity in general—was the dominant selective force. Notably, most of the shift in the pI distribution occurred in the latter half of the experiment (Fig. 3E), highlighting the importance of sustained rapid mutagenesis over long periods of evolution to embed such presumably subtle selective forces into the data.

TrpB's pI evolved to be comfortably below the typical yeast cytosolic pH of 6.8 to 7.2 (*62*), which is consistent with the notion that intracellular proteins prefer to be negatively charged to minimize large-scale clustering with RNAs and other proteins (*49*, *63*). One possible mechanism by

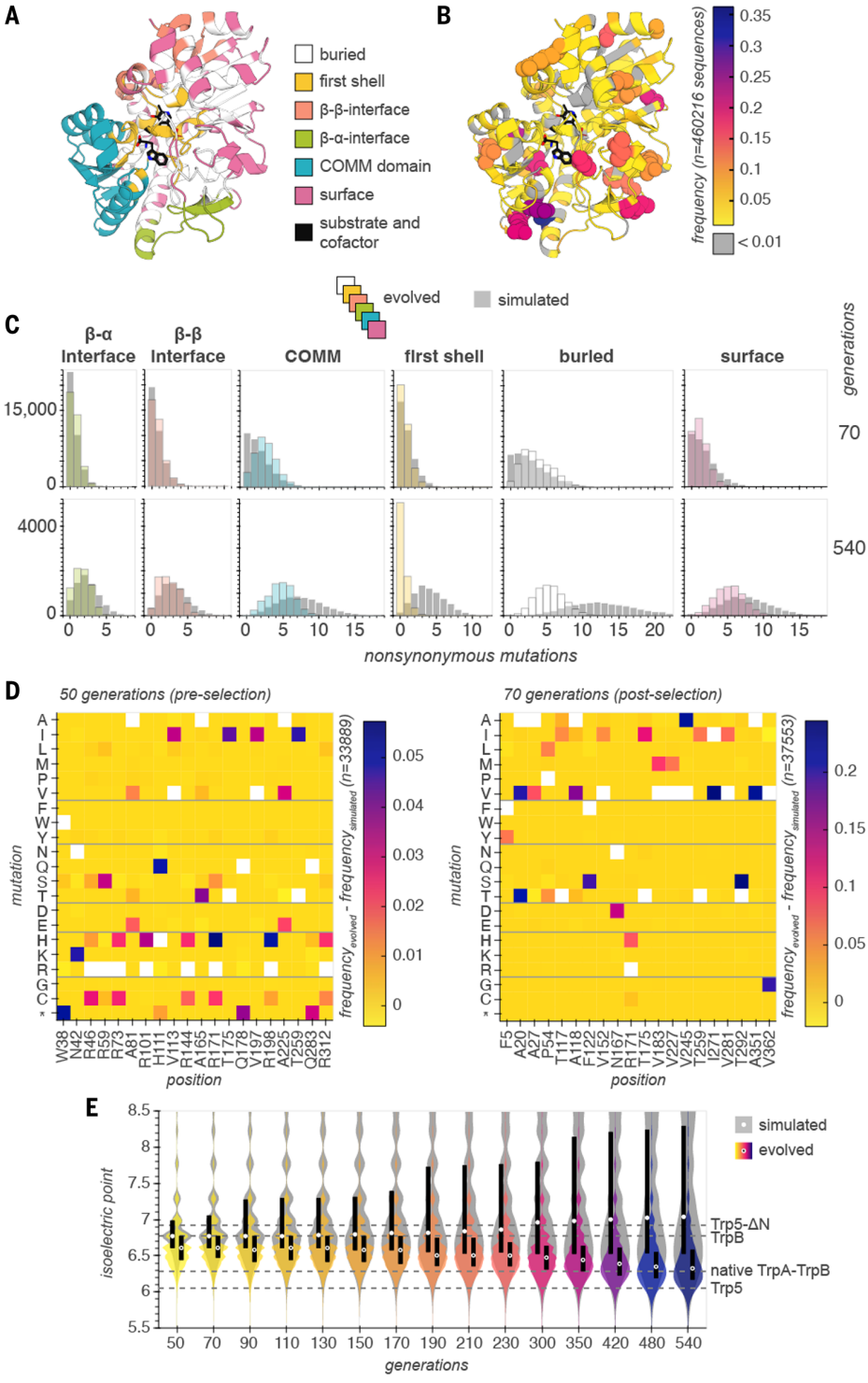


Fig. 3. Revealed effects of selective constraints. (**A**) AlphaFold structure of *T. maritima* TrpB with different regions colored according to their structural role. First shell, β-β interface, and β-α interface residues are designated as such if they are within 5 Å of the substrate and cofactor (Trp and PLP), the other β subunit, or the other α subunit in the αββα heterotetramer holoenzyme, respectively. Alignment to *Pyrococcus furiosus* TrpB crystal structures (PDB codes 5EOK and 5DW3) was used to determine distances from substrate and cofactor, α subunit, and β subunit. Mean solvent accessible surface area (SASA) was used to categorize all remaining residues as either surface (SASA ≥ 0.2) or buried (SASA < 0.2). (**B**) Heatmap of mutations among OrthoRep-evolved TrpB sequences applied to the AlphaFold structure. (**C**) Distributions of mutations among OrthoRep-evolved TrpB sequences within the six structural regions compared with a null model composed of a simulated dataset of sequences mutagenized in silico according to the mutation rates and preferences measured for TP-DNAP1 BadBoy2 until the number of synonymous mutations in the simulated sequences and real sequences were equivalent. (**D**) Heatmaps of mutation frequency for all mutations among the 20 most frequently mutated positions in the time point corresponding to either 50 (left) or 70 (right) generations. Frequencies for simulated sequences were subtracted to account for bias due to BadBoy2's mutation preference and WT TrpB sequence content. (**E**) Violin plots of isoelectric points for all OrthoRep-evolved and simulated TrpB sequences, split by time point. Points and black bars denote the means and interquartile range for all sequences within each time point. Isoelectric points of the WT *T. maritima* TrpB, an N-terminally truncated Trp5 homologous to TrpB (Trp5-ΔN), the native *T. maritima* TrpA-TrpB complex, and the TrpA-TrpB holoenzyme ortholog from *S. cerevisiae* (Trp5) are shown for comparison. Note that the TrpB sequence used as a starting point for evolution includes a 6xHis tag that contributes an increase in pI of 0.2 and that the native TrpA-TrpB complex pI shown for reference does not include this 6xHis tag.

which this preference could have driven the observed adaptation is through its influence on TrpB function itself—for example, by increasing the diffusivity of the enzyme (*64*). Another mechanism by which a preference for negative charge in TrpB could have been adaptive is by lessening its perturbation on other entities in the cell—for example, by preventing spurious association or aggregation that would disturb the function of the proteome (*49*). Our data do not exclude either mechanism but suggest that the latter mechanism is present: In generations 0 to 50 of the evolution experiment, TrpB pIs were significantly lower than those of the null model sequences ($P < 0.0001$) even though TrpB was not under selection for function as excess Trp was supplied to the growth media. Consistent with the idea that WT TrpB's pI is disruptive to the cell and affords an opportunity for adaptation separate from enzymatic function, we also found that stop codons were among the most enriched mutations (Fig. 3D) over the null model after the first 50 generations. Our observation of charge optimization even when there was no selection for the enzyme's function demonstrates that the intracellular context can impose constraints on

the physicochemical properties of proteins independent of its primary molecular function. It also highlights the value of evolving proteins in vivo, where subtle constraints dictating intracellular compatibility can be both revealed and included in the evolutionary optimization of protein function.

Thermoadaptation

Given that our parental *T. maritima* TrpB was from a thermophile but needed to evolve standalone activity in a mesophile, we looked for statistical evidence of thermoadaptation in our evolved sequences. Haney *et al*. have studied the patterns of amino acid replacements between natural orthologous proteins in mesophilic versus thermophilic organisms and have found 17 amino acid replacements that distinguish the mesophilic variants from the thermophilic variants at homologous positions in multiple sequence alignments with high confidence (*13*). When we evaluated the frequency of these 17 amino acid replacements among all mutations in our evolution experiment's outcomes, we found that replacements in the mesophilic direction were enriched (fig. S20). As before, this illustrates the ability of extensive gene evolution to reveal selective forces through the evolutionary information embedded into the resulting diversity.

Networks of coupled mutations

In addition to global properties, we investigated finer patterns in the outcomes of TrpB evolution with the expectation that these may reveal coevolving networks of amino acids responsible for adaptation and divergence. At the beginning of our evolution experiment, we had included short barcodes adjacent to the TrpB sequence integrated onto p1. This allowed us to (i) isolate the largest clades for analysis, because these should correspond to the fittest sequences, and (ii) reduce the contribution of phylogeny by computationally limiting the number of sequences analyzed per clade (Fig. 4A). The latter increases the signature of mutations independently discovered across multiple clades, favoring the detection of mutations whose cooccurrence was functionally important. Specifically, we considered clades with barcodes that had at least 100 associated sequences—there were 93 such clades—and randomly downsampled to 100 sequences for clades whose members exceeded this number.

This analysis revealed that some sets of mutations frequently cooccur (Fig. 4B). In one particular example, the A20V (Ala^{20}→Val) mutation was found to cooccur at a frequency above ~0.8 with a set of four auxiliary mutations (A118V, F122S, I271V, and T292S) in eight distinct lineages in later time points, which implies a strong relationship among these mutations (Fig. 4, C and D). This set of mutations, as well as other sets identified, are spread throughout the structure of TrpB, in line with recent studies demonstrating the prevalence of structurally distributed allosterically activating mutations in TrpB and other proteins (*6*, *60*, *61*, *65*). Among these four auxiliary mutations is the T292S mutation, which was previously identified in a TrpB directed evolution campaign as highly activating, alone conferring a more than fivefold increase in the standalone catalytic efficiency of TrpB (*20*). The association among mutations was sensitive to their specific identities. For example, the A20T mutation was individually present at a higher frequency than A20V among all sequences (fig. S16D) but was not associated with any specific other mutation at a frequency above ~0.4 (Fig. 4D). This suggested that the A20T mutation might be broadly activating across more sequence contexts than the A20V mutation, implying long-range epistatic interactions, which we partially investigated (supplementary text and fig. S21).

Fitness of TrpBs and their predictability

Accurately modeling the fitness landscapes of proteins is a major goal of ML. However, it is known that ML models are biased toward favoring sequences that are more similar to the natural sequences on which they were trained (*66*), and it remains unclear to what extent they can predict the function of sequences that are many mutations away from these natural sequences. It is also unclear whether ML can model how mutant sequences will perform on new functions that deviate from natural functions in service of bioengineering goals, such as enzyme and antibody engineering. To provide insight into these questions, we tested whether an advanced ML model could predict the fitness of our evolutionary outcomes.

To gain high-resolution fitness information on evolved TrpBs, we first profiled evolved variants in a high-throughput enrichment assay (Fig. 5A). We cloned a library of ~100,000 TrpB sequences isolated from our evolution experiment into a standard yeast plasmid that would not be subject to hypermutation by OrthoRep, transformed this library into yeast, applied selection for TrpB function, and tracked the enrichment or depletion of individual variants using HTS. Included in this library were two previously engineered control TrpB variants known to be either highly functional (TrpB-003-1-A) or nearly nonfunctional (*Tm*Triple) in the context of yeast Trp production growth complementation (Fig. 5B) (*38*). We evaluated Trp production by members of this library using three distinct growth conditions: Trp-supplemented media (no selection), media lacking Trp with a high concentration of indole (400 μM, weak selection), and media lacking Trp with a low concentration of indole (25 μM, strong selection). We tracked the abundance of library members in replicate yeast transformations of the same library over four time points taken at the beginning and end of six passages to obtain fitness scores. We found that fitness scores above a threshold (enrichment score > −5) were well correlated among replicates for both weak and strong selection conditions (Pearson correlation = 0.79 and 0.84, respectively) but not for nonselective conditions where Trp was present (Fig. 5C), which confirms the reliability of the assay for highly functional TrpB variants. Thousands of multimutation sequences at least as functional as the previously engineered high-fitness TrpB-003-1-A, with a k_{cat}/K_M of $1.4 \times 10^5\ M^{-1}\ s^{-1}$ (*38*), were identified (Fig. 5D). We also observed that a large fraction of sequences had low activity (Fig. 5D), which likely owes to the multicopy nature of p1 (fig. S13) that creates a delay in the action of purifying selection on recently generated mutants hitchhiking with functional TrpBs in the same cell. Although such low-activity sequences may be of little direct use, we note that they can contain negative design information of value.

We then investigated whether a state-of-the-art ML model called TranceptEVE (*67*), which ensembles an autoregressive large language model (Tranception) trained across protein families with a variational autoencoder (EVE) trained on a specific family of proteins (in this case, TrpBs), could predict the measured fitness scores of our laboratory-evolved TrpBs (Fig. 5, E and F). Although sequences that were predicted to have low fitness did exhibit little or no function in our enrichment assay, we found essentially no correlation between the predicted scores and the real enrichment scores of high-function TrpBs (Fig. 5F; strong selection mean enrichment score > −5; Pearson correlation, 0.033; $P < 0.0001$). For example, the highest predicted score was assigned to the nearly nonfunctional *Tm*Triple variant. By contrast and as expected, we found that predicted scores exhibited a much stronger and negative correlation with the number of nonsynonymous amino acid mutations (Pearson correlation, −0.765; $P < 0.0001$; Fig. 5F). One explanation for the low predictability of our evolved TrpBs' relative activities is that the TrpB standalone function they gained is rarely high in the natural sequences on which TranceptEVE was trained. Another explanation is that drift under purifying selection brought our evolutionary outcomes into regions that are out of distribution of natural sequences simply due to the extent of novel divergence achieved in our experiment. Understanding the relative contribution of these two explanations will require future work, including OrthoRep experiments that only challenge genes to maintain their existing function without the need to adapt. Nevertheless, the ability for scalable continuous evolution to enter and explore functional regions of fitness landscapes that ML models do not, and vice-versa

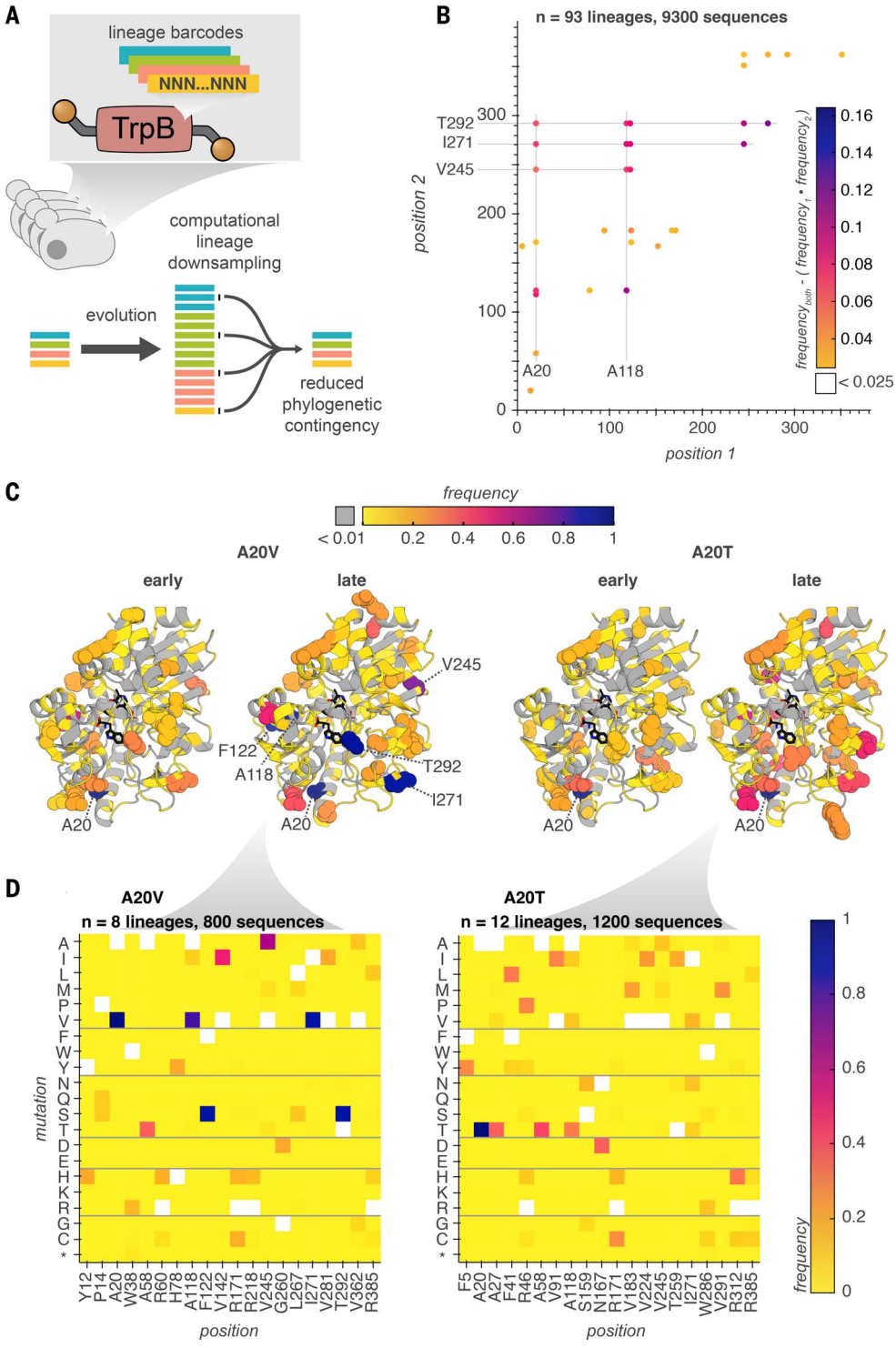


Fig. 4. Lineage barcodes reveal covarying residues. (**A**) Schematic of computational processing used to reduce phylogenetic contingency of residue covariation. (**B**) Residue covariation plot for the most frequently covarying residues among all time points in the TrpB evolution dataset for 93 lineages downsampled to 100 sequences per lineage. (**C** and **D**) Heatmaps of most frequently mutated residues among sequences containing mutations A20V or A20T, downsampled to 100 sequences and chosen from specific time points. Heatmaps of all positions for early (generations 70 and 90) and late (generations 480 and 540) time points are overlaid onto a TrpB AlphaFold structure with the 20 most frequently mutated positions shown as spheres (C) or shown as mutation-specific heatmaps of the most frequently mutated 20 positions in late time points (D).

(*31*, *68*), highlights both open challenges in ML and the potential value of combining these two types of approaches going forward.

Discussion

In this work, we have engineered OrthoRep's mutation rate to reach >10^{-4} s.p.b. while also reducing its bias against transversion mutations. These upgraded OrthoRep systems successively drove the extensive sequence divergence of a chosen gene, TrpB, through positive selection regimes, where new adaptations emerged; no selection regimes, where sequences aimlessly diversified; and purifying selection regimes, where high mutation rates were critical in revealing patterns of conservation and change from which we could infer structural and functional constraints. Although TrpB was used to demonstrate the capabilities of OrthoRep in both the evolutionary improvement of a gene's function and the evolutionary recording of selective forces into sequence diversity, our experiment should readily extend to any selectable gene. Our experiment should also be capable of scaling beyond 96 replicate lines, including through assistance by automated liquid handling systems (*69*, *70*),

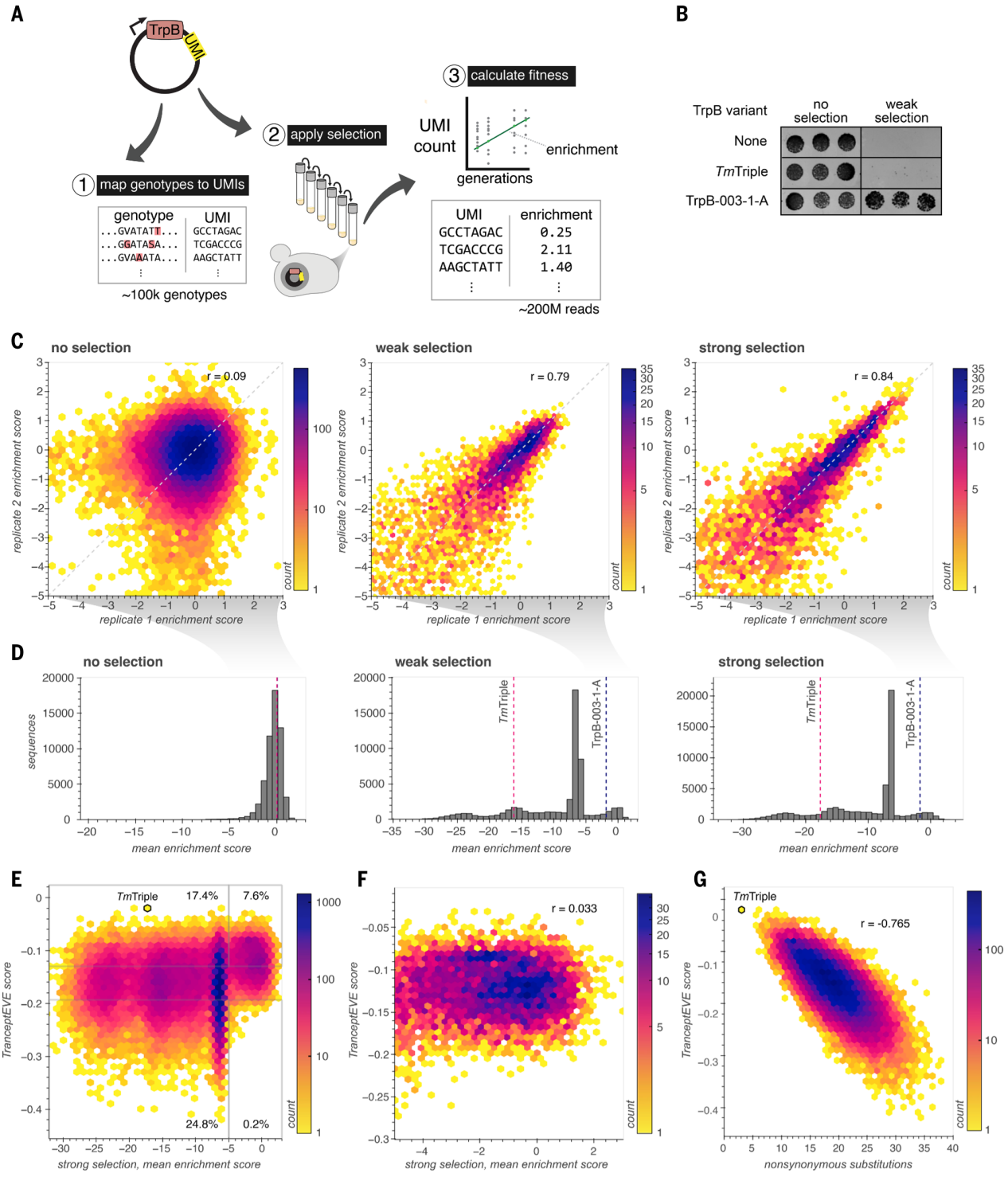


Fig. 5. Pooled measurement and TranceptEVE prediction of TrpB variant fitness. (**A**) Schematic of pooled TrpB fitness assay using HTS. (**B**) Spot plating growth assay of control sequences included in the pooled fitness assay. (**C**) Hexbin plots of replicate concordance among pairs of replicates under growth conditions with Trp (no selection), without Trp and with 400 uM indole (weak selection), or without Trp and with 25 uM indole (strong selection) for highly functional sequences (enrichment score > −5). (**D**) Distributions of mean enrichment scores (average of *n* = 2 biological replicates) among the three selection conditions. (**E** and **F**) Hexbin plots of TranceptEVE score versus measured mean enrichment score with strong selection for either all enrichment scores (E) or enrichment scores for highly active sequences (F). The percentage of all sequences that fall in the upper or lower quartile of score predictions and are classified as either high or low function (enrichment score greater than or less than −5, respectively) are shown in the respective sections of the plot in (E). (**G**) Hexbin plot of TranceptEVE score versus number of nonsynonymous substitutions for all sequences with a measured strong selection mean enrichment score. r, Pearson correlation.

to support the generation and maintenance of greater gene diversity, to detect independently discovered covarying networks of mutations with greater power, and to carry out comparative evolution experiments across different selection schedules. Overall, the practicality of condensing long adaptive and neutral gene evolutionary processes into laboratory experiments realized in this work should find broad applications in the evolutionary engineering of biomolecules, in the finer and broader mapping of sequence-function relationships, in revealing previously unknown biological constraints that shape evolution, and in understanding how genes evolve—from their own points of view.

Materials and methods

DNA plasmid construction

Plasmids used in this study are listed in table S4, along with sources for DNA templates. Complete maps for these plasmids are available on Zenodo (*71*). All DNA templates for polymerase chain reaction (PCR) were derived from previous studies or gBlocks (IDT). All primers were synthesized by IDT. All relevant primer pairs are listed in table S5. Amplicons for construction of clonal plasmids were generated using Q5 Hot Start High-Fidelity DNA Polymerase (NEB). All nonlibrary plasmids were constructed using Gibson Assembly and transformed into chemically competent *Escherichia coli* strain TOP10 (ThermoFisher). Clonal plasmids were sequence verified by either Sanger sequencing (Azenta) or whole-plasmid sequencing (Primordium).

DNA library construction

Amplicons for TP-DNAP1 libraries were generated using error-prone PCR with GeneMorph II (Agilent) according to manufacturer instructions, aiming for ~3 to 5 nucleotide substitutions per sequence. Amplicons for all other libraries were generated with Q5 Hot Start High-Fidelity DNA Polymerase (NEB). For epPCR 1, the resulting PCR product was assembled into plasmids using Gibson assembly in 20 μl reaction volumes. For all other libraries, resulting PCR products were assembled into plasmids with Golden Gate assembly with T4 DNA ligase and BsaI-HF v2 or PaqCI (all NEB) in a 40 μl reaction volume. Gibson reactions were run at 50°C for 1 hour. Golden gate reactions were run isothermally at 37°C for 1 hour and heat inactivated at 65°C for 10 min. Reactions were purified with AMPure XP beads (Beckman), typically with a 0.9:1 bead:sample ratio according to manufacturer instructions. Libraries were transformed into high-competency electrocompetent *E. coli* TOP10 cells (ThermoFisher).

Yeast strains, media, transformations, and DNA extraction

All yeast strains used in this study and their provenance are listed in table S6. Yeast were grown in liquid or on plates at 30°C in synthetic complete (SC) growth medium [20 g/liter dextrose, 6.7 g/liter yeast nitrogen base with ammonium sulfate without amino acids (US Biological), appropriate nutrient drop-out mix (US Biological), as directed] or MSG SC growth medium [20 g/liter dextrose, 1.72 g/liter yeast nitrogen base without ammonium sulfate without amino acids (US Biological), appropriate nutrient drop-out mix (US Biological), as directed, 1 g/liter L-Glutamic acid monosodium salt hydrate (ThermoFisher)] minus nutrients (referred to as -X where X is either the single letter amino acid code for an amino acid nutrient, or U for uracil) required for appropriate auxotrophy selection(s). Where selection for MET15 was required, cells were propagated in media lacking both methionine and cysteine. 500 μl liquid yeast cultures in 96-well deep well plates were incubated with shaking at 750 rpm. All other liquid yeast cultures were incubated with shaking at 200 rpm.

Yeast transformations, including p1 integrations and polymerase replacement integrations, were performed as previously described (*38*). For all integration transformations, plasmid DNA was linearized before transformation using either ScaI-HF or EcoRI-HF (both NEB) for p1 or genomic integrations, respectively. Due to its repetitive nature, deletion of FLO1 was performed by a URA3 knock-in knock-out method (*72*) (table S4, pFLO1-KO). Genetic deletions for TRP5 and MET15 were performed as previously described (*73*), using spacer sequences TTTGAGCCTGATCCCACTAG and GCTAAGAAGTATCTATCTAA, respectively. When isolating individual clones from genetic deletion and integration transformations, colonies were restreaked onto media agar plates of the same formulation to ensure isolation of only cells that have the desired genetic change.

All p1 plasmid sequences were generated by first generating a strain harboring a "landing pad" p1 via integration and then integrating over this landing pad to generate the desired p1 construct. To enable construction of the landing pad strain, the WT TP-DNAP1 was integrated at the CAN1 locus using pGR475. A sequence encoding a nonfunctional partial LEU2 sequence lacking the N terminus was then integrated over the WT p1 using pGR420 to generate the landing pad p1. The WT p1, which encodes the TP-DNAP1, was then cured out via three to four 1:1000 passages. The p1 plasmid(s) encoding the desired sequence were then generated via integration using cassette(s) that include a LEU2 sequence lacking the C terminus (e.g., pGR438). The overlap between the LEU2 on the landing pad and the new integration cassette reconstitutes full length LEU2 only when integration occurs on p1, reducing the likelihood of genomic integration.

The polymerase replacement integration transformation was performed by first digesting 0.5 to 2 μg of the polymerase replacement plasmid or library with EcoRI-HF in a 25 μl reaction per 1x transformation followed by directly transforming this digestion reaction into a yeast strain encoding the CAN1-WT-TP-DNAP1 landing pad (all polymerase libraries were transformed into OR-Y488). Library transformations were carried out at 20 to 40x scale. Transformed yeast were plated onto solid MSG SC -LR or -MCR media with 100 mg/liter nourseothricin (for positive selection of integration) and 200 mg/liter L-canavanine (a toxic L-arginine analog for counterselection of cells that fail to perform polymerase replacement and remove the arginine permease CAN1). Leu or Met/Cys dropout was used to maintain selection for p1-encoded LEU2 or MET15, respectively, whereas Arg dropout was used to improve L-canavanine selection.

Extraction of genomic DNA (gDNA) and p1/p2 plasmids was performed as previously described for 1.5 ml yeast culture volumes (*38*). This procedure was used for all experiments except for DNA extracted for use in HTS dataset 6, which was instead performed in 96-well format. In brief, a 96-well block of 500 μl of saturated yeast cultures was centrifuged (2500 ×*g*, 5 min), supernatant was discarded, pellets were resuspended in 1 ml 0.9% NaCl, this resuspension was again centrifuged (2500 ×*g*, 5 min), and the supernatant was discarded. The resulting pellet was resuspended in 250 μl Zymolyase solution [0.9 M D-Sorbitol (Sigma Aldrich), 0.1 M Ethylenediaminetetraacetic acid (EDTA) (Sigma Aldrich), 10 U/ml Zymolyase (US Biological)] and incubated with shaking (37°C, 200 rpm). The 96-well block was then centrifuged (2500 ×*g*, 5 min), supernatant was discarded, and pellets were resuspended in 280.5 μl of proteinase K solution {250 μl TE [50 mM Tris-HCl (pH 7.5), 20 mM EDTA], 25 μl 10% sodium dodecyl sulfate (SDS) (Sigma Aldrich), 5.5 μl proteinase K stock solution (10 mg/ml proteinase K (ThermoFisher)}. The 96-well block was then incubated at 65°C for 30 min, combined with 75 μl of 5M potassium acetate (ThermoFisher), and incubated on ice for 30 min. The 96-well block was centrifuged at 12,000 ×*g* for 10 min, the resulting supernatant was combined and mixed with 2 volumes buffer PB [5 M Guanidine hydrochloride (ThermoFisher), 30% isopropanol, 70% water], and this mixture was applied to a 96-well DNA-binding plate (Epoch Life Science) on a vacuum manifold. Flow through was discarded, columns were washed with PE buffer [10 mM Tris-HCl (ThermoFisher), 80% ethanol, 20% water, pH 7.5], centrifuged and dried, and 60 μl of water was applied to columns for elution by centrifugation (2500 ×*g*, 5 min).

HTS

All HTS datasets are listed in table S7, along with the method used to construct them. All

PCRs for HTS were performed with Platinum SuperFi II DNA Polymerase (ThermoFisher). For short-read paired-end sequencing, both low-yield (AmpliconEZ, Azenta) and high-yield (HiSeq paired end 150, Novogene) were performed directly on PCR products generated in either one or two rounds of PCR using primers that each included an adapter sequence, a 6- or 7-nucleotide barcode, or both.

For in-house long-read HTS, we used the ONT nanopore sequencing platform. Due to the lower accuracy of nanopore sequencing, we used modified versions of previously described methods for the construction of DNA libraries that yield multiple reads of the same original DNA molecule, allowing for computational reconstruction of high-accuracy sequences (*53*, *73*). We refer to the first of these as in vivo downsampled UMI PCR, which was used for most nanopore sequencing in this study (table S7). This involved first a 2-cycle "UMI tagging" reaction, in which primers (e.g., primer pair 2) were used to append both UMIs and universal DNA sequences (for further amplification) to both ends of the target sequence with the following components: (i) ~1 to 50 ng of purified yeast or *E. coli* miniprep, (ii) 5 μl of 2x SuperFi II master mix, (iii) 1 μM each primer, and (iv) water to 10 μl; and with these thermocycler conditions: (step 1) 98°C for 30 s; (step 2) 98°C for 10 s; (step 3) 65°C for 1 s; (step 4) 60°C for 45 s with ramp down from 65 to 60 at 0.2°C per second (this should result in 25 s of ramp down time and 20 s of hold time); (step 5) 72°C extension, 1 min/kb; (step 6) go to step 2 (1x); (step 7) 72°C for 2 min.

Next, UMI primers were removed using ExoSAP-IT (ThermoFisher) according to manufacturer instructions. 10 μl of the resulting reaction was then used as template in a 25 μl Platinum SuperFi II PCR with 1 mM $MgCl_2$ supplemented, using primers (e.g., primer pair 3) that included BsaI or PaqCI sites and seven nucleotide barcodes in forward-reverse combinations that were unique for each sample. The resulting uniquely barcoded PCR products were then combined, purified using AMPure XP beads, used in a library Golden Gate reaction with an *E. coli* vector, then transformed into high competency *E. coli* as described above. Plasmid pGR554 (table S4) was designed for this purpose and contains CcdB and sfGFP, which both are replaced with the insert during Golden Gate assembly, as well as NotI and SbfI sites, strategically placed to enable separation of the desired library insert from the backbone before sequencing. CcdB and sfGFP provided counterselection and visualization of colonies resulting from undigested vector. [Cloning into any *E. coli* vector will suffice however, so long as unique library members are associated with a unique relatively short (20 to 50 base pair) sequence.] Resulting colonies each contained many copies of a unique plasmid species encoding a UMI-tagged library member. To obtain good coverage of each sequence and UMI with nanopore sequencing, the resulting library was downsampled by only harvesting ~20-fold fewer colonies than the expected number of reads, amounting to 100 to 200 thousand colonies for a standard MinION flow cell. Plasmid DNA from this library was then miniprepped, digested [for example with NotI-HF or NotI-HF/SbfI-HF (both NEB) if using plasmid pGR554], and gel extracted before sequencing.

Construction of the library for HTS dataset 8a was performed using a similar approach, but with a yeast expression vector, and with UMIs and library members inserted into the vector in two distinct steps so that UMIs are present at a single location in the plasmid and would not need to be immediately adjacent to library members, mitigating any potential effects of UMIs on expression. In brief, libraries of UMIs generated via PCR using primer pair 8 were cloned into plasmid pUMI by Golden Gate assembly with BsmBI-v2 and *E. coli* electrotransformation to generate an intermediate UMI library. Different variants of pUMIs with unique, known 7-nucleotide barcodes were used for each library to allow multiplexing. These libraries were then used as the vector into which evolved TrpB library members, PCR amplified using primer pair 9, were cloned via standard restriction cloning, with inserts digested with PaqCI and vectors digested with BsaI-HFv2. Isothermal ligation (1 hour 37°C, 20 min 65°C) was then performed with T4 ligase in 40 μl reaction volumes, AMPure bead purified, and transformed into high-efficiency *E. coli*. This was carried out for six unique libraries: downsampled in yeast with selection, downsampled in yeast without selection, and not downsampled in yeast, each for both time points (generation 350 and generation 540) chosen from TrpB evolution. The intermediate UMI libraries were generated at a library size of >100-fold larger than the desired final library size to minimize the chance that distinct library members would be tagged with identical UMIs.

The second method used for generating libraries for high-accuracy nanopore sequencing was adapted from methods described in Volden *et al.* (*74*), Oliynyk and Church (*75*), and Zhang and Tanner (*76*). It involved circularization of the target sequence and use of this circularized product as template for rolling circle amplification (RCA) using strand displacing DNA polymerases (fig. S15). In brief, UMI-tagged PCR products were generated as described above, albeit with complementary Type IIS cut sites (BsaI or PaqCI) on the forward and reverse primers used during PCR amplification such that the two ends of the amplicon ligated to each other during a Golden Gate assembly reaction, forming a circular product. Due to the large amount of template DNA required for RCA, a large amount of amplicon was used in the circularization reaction, typically 1 to 2 pmols in 100 μl Golden Gate assembly reactions. Oliynyk and Church (*75*) provide a useful discussion of relevant considerations for such circularization reactions. An isothermal Golden Gate assembly reaction was then performed to circularize the amplicon library.

After Golden Gate assembly, uncircularized DNA was digested with lambda exonuclease (NEB), exonuclease I (NEB), and exonuclease III (NEB) in a 10:10:1 ratio, which was added directly to the Golden Gate reaction in a 1:10 exonuclease mixture to sample ratio. This reaction was incubated at 37°C for 45 min, then 80°C for 15 min. The reaction was then AMPure bead purified with a 0.7:1 bead:sample ratio, eluting in a maximum of 10 μl of water, and the resulting circularized library was used in a RCA reaction with the following components combined on ice: (i) 4 μl 10X NEB buffer 4 (NEB); (ii) 4.8 μl 10 mM dNTPs; (iii) 2.64 μl 5 U/μl Bsu DNA Polymerase, Large Fragment (NEB); (iv) 1.6 μl 10 μg/μl T4 gene 32 protein (NEB); (v) 0.2 to 1 μg circularized DNA library; (vi) RCA primers, 2 μM each (must bind internal to first primer set, e.g., primer pair 7); and (vii) water to 40 μl.

This reaction was incubated at 37°C for 3 hours. SDS-containing loading dye was added directly to the reaction, and the entire sample was run on an agarose gel. Bands corresponding to 3x to 6x concatemers were gel extracted, and purified DNA was used for nanopore sequencing. Unlike RCA using Phi29, this method enabled size selection of specific repeat numbers and did not require a "debranching" step but required large amounts of input DNA and in our experience suffered from high sensitivity to DNA contamination. Use of Phi29 with random hexamers, followed by debranching, is therefore a reasonable alternative.

After construction of in vivo downsampled UMI PCR or RCA libraries, nanopore library preparation and sequencing was performed using the most up-to-date ligation sequencing kit (e.g. LSK-114) and flow cell (e.g. R10.4.1), following manufacturer instructions, with two exceptions. First, ½ volumes (but unaltered DNA input) for end prep and ligation reactions were used and second, FFPE Repair Mix was not used during end prep reactions.

All relevant HTS datasets are made available on the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA), accession no. PRJNA1050257.

Error rate measurement by mutation accumulation

After a polymerase replacement integration transformation, colonies were picked into liquid

media of the same media formulation as that used for selection after transformation, then grown to saturation. The resulting saturated culture was miniprepped to serve as the 0th passage, p_0. The culture was then also propagated in the appropriate growth medium for maintenance of the orthogonal plasmid, typically SC -L, using a dilution factor d (typically 128, 256, or 512) that is consistent throughout the experiment. At least one additional saturated culture from this time course was miniprepped to serve as the t^{th} passage, p_t. We used only two passages/time points in all mutation accumulation error rate measurement experiments except that which produced HTS dataset 6. The number of generations, g, that separated p_0 from p_t was used to calculate the mutation rate and can be approximated as

$$g_t = t \times \log_2 d$$

We note that this approximation assumes no cell death and equivalent saturation at each passage. Cultures were manually propagated several times until a total number of generations of at least 50 was reached. Miniprepped yeast DNA for each time point was used as template DNA for HTS, and custom scripts, organized within the Maple pipeline, were used to calculate mutation rates.

Mutation rates were calculated as the rate of accumulation of a mutation type (all substitutions, individual substitutions, insertions, or deletions), j. We denote μ_j as this rate calculated for mutation type j. High-accuracy HTS was used to first obtain $c_{j,t}$, the total counts of mutation j (e.g., A to T) among all sequences in passage t. For substitution mutations, to account for the influence of variable A/T/G/C content in the sequence being analyzed, this count is normalized to obtain $n_{j,t}$, the expected count for an idealized sequence with a 1:1:1:1 A:T:G:C ratio. For a substitution at a particular nucleotide where the nucleotide occurs w times within a reference sequence of length L, $n_{j,t}$ is calculated as

$$n_{j,t} = \frac{c_{j,t}}{w} \times \frac{L}{4}$$

The total normalized count of all substitution mutations at each time point is then calculated as the sum of all $n_{j,t}$ for all 12 substitution types. However, for insertions and deletions, $n_{j,t}$ is not normalized in this way, and is instead equivalent to $c_{j,t}$, the total number of nucleotides inserted or deleted among all sequences for that time point.

To obtain the per-nucleotide, per-generation mutation rate μ_j, we used the total number of sequences analyzed for each time point, s_t, to calculate the per-nucleotide frequency of mutation j for each time point. We then use linear regression on these normalized per-nucleotide frequencies according to

$$\frac{n_{j,t}}{L \times s_t} = \mu_j \times g_t + b_j$$

where μ_j and b_j are the slope and intercept, respectively, of the best fit line for all t time points. When the number of generations between the 0th time point and the initiation of mutagenesis (typically polymerase replacement) is accurately estimated and no mutations fully fixed within the population before the first time point, b_j should be close to 0. Regardless, we do not report b_j, as it has no bearing on μ_j across experiments. We report μ_j as the per-base per-generation rate of accumulation of mutation type j. Mutation tabulation was performed by the script mutation_analysis.py and all other operations related to mutation rate calculation were performed by the script plot_mutation_rates.py, both of which are contained within the Maple pipeline. All reported mutation rates were calculated using mutation tabulation within a sequence region that was not under functional selection.

TP-DNAP1 library selection and screening

Error-prone TP-DNAP1 libraries were cloned in *E. coli*. Mutagenesis was validated by Sanger sequencing of 8 to 12 individual clones. A summary of these libraries can be found in table S8.

After TP-DNAP1 replacement library construction, resulting purified plasmid DNA was used for a polymerase replacement integration transformation. Before transformation, OR-Y488 was grown up in SC -L + 1 mg/liter 5-fluoroorotic acid (US Biological) for counterselection against cells that had reverted the inactivating mutation in ura3* by chance. After library transformation, plating on media selecting for cells that had replaced the WT TP-DNAP1 with library variants, and 48-hour incubation, colonies were harvested in bulk and immediately plated onto SC -LU plates. These plates were then incubated for 48 hours, and resulting colonies were either picked into liquid media for mutation rate or frequency characterization (by mutation accumulation or fluctuation assays, respectively) or were harvested in bulk and immediately plated onto SC -LUW media. After colony formation, colonies were either picked into liquid media for mutation rate characterization by mutation accumulation or were harvested in bulk, miniprepped for gDNA isolation, and used as template for a nonmutagenic PCR to generate amplicons for Golden Gate assembly into pGR554, which was then retransformed into OR-Y488 to repeat the selection and perform mutation rate screening.

The fluctuation test was performed as follows. After transformation of the epPCR 1 TP-DNAP1 library into OR-Y488 and selection for ura3* reversion on solid media, individual colonies were picked from this plate, inoculated into 500 µl SC -LU media in a 96-well block, and grown to saturation. Cultures were then passaged 1:10,000 into 200 µl SC -LU, 12 replicates per each individual colony, and grown to saturation. Cultures were centrifuged, washed with 0.9% NaCl, and pellets were resuspended in 35 µl 0.9% NaCl. 10 µl of each resuspension was then plated onto SC -LUW plates. A subset of cultures were titered and plated on SC -LU plates to estimate population size. Plated cells were allowed to grow for 4 days, and revertants on each spot were counted. Counts were used to estimate the m value using the FALCOR online web tool (https://lianglab.brocku.ca/FALCOR/) and the Ma-Sandri-Sarkar Maximum Likelihood Estimator. Mutation frequency was calculated from this m value as previously described (*42*), using a target size of 1 (only one mutation is capable of restoring Trp5 activity). Copy number was not considered and therefore per-base substitution rate was not calculated using this method.

Copy number measurement

Quantitative PCR (qPCR) was used to determine p1 copy number. gDNA was isolated from samples according to the 1.5 ml DNA extraction protocol referred to above. qPCR reactions were performed in 10 µl volumes using the Sybr Powerup Master Mix (ThermoFisher), with 1 to 5 ng of gDNA template. Reactions were run at "standard speed" on a ThermoFisher Quantstudio 6.

The reactions measured the amplification of genomic *GAL1*, and p1-encoded *LEU2* using primer pairs 16 and 17, respectively (table S5), yielding cycle threshold (C_t) values for each sample. A standard curve was generated correlating C_t with DNA quantity using reactions containing known quantities of plasmids encoding *GAL1* and *LEU2*. Standard curves were used to validate the assumption that the two primer pairs had the same amplification efficiency. Relative copy numbers of *GAL1* and *LEU2* in each sample were determined from the standard curves, and absolute p1 copy numbers were normalized to genomic copy numbers to obtain absolute per-cell p1 copy number by dividing *LEU2* relative copy numbers by *GAL1* relative copy numbers for each sample.

TrpB evolution

Plasmid pGR595 (TP-DNAP1 BadBoy2, promoter SAC6) was first transformed into yeast strain OR-Y484 according to the polymerase replacement procedure described above. The resulting strain (OR-Y538) was then transformed with plasmid pGR438 (TrpB with lineage barcodes), following the p1 integration procedure described above, plating on SC -L media. ~400 resulting colonies were harvested together and passaged into 512 µl SC -L media in all wells of a 96-well block, grown to saturation, then again passaged 1:1024 into SC -L media. DNA extracted from these resulting cultures served as passage/time point 0, which we approximate

to be ~50 generations from TrpB p1 integration. These cultures were also passaged 1:1024 (0.5 μl into 512 μl) for all passages in the experiment into growth media and with time points taken as described in table S2. DNA extraction for time points was performed by combining all 96 saturated cultures for a specific time point and extracting DNA from the pooled cultures according to the 1.5 ml DNA extraction protocol referred to above.

Selection pressures shown in table S2 and Fig. 2B were derived from the concentrations of Trp and indole in the growth media used for each passage. These two components are inversely related to the selection pressure supplied by each component, scaled to range from 0 to 0.5 such that the maximum concentration used for both Trp and indole would yield a selection pressure of 0, media without Trp and with the maximum concentration of indole would yield a selection pressure of 0.5, and media without either component would yield a selection pressure of 1. The three selection periods no selection, mostly positive, and mostly purifying were characterized as such based on the selection pressure used during that period. The initial no selection period only included the minimum selection pressure (0). The subsequent mostly positive period proceeded until the populations were exposed to the highest selection pressure for the first time. Subsequently, populations did not need to further adapt to more stringent selection conditions, so we refer to the following period as mostly purifying.

To downsample evolved populations for cloning UMI-tagged TrpB libraries, yeast cultures from the passages corresponding to generation 340 and 510 were inoculated into SC -L media from glycerol stock and grown to saturation. Saturated cultures were then combined, and multiple serial dilutions of both cultures were plated onto SC -L media. Plates derived from generation 340 and 510 with ~3700 and ~1700 colonies, respectively, were harvested. These cultures were passaged 1:1000 into 2 ml of either SC -LW + 400 μM indole (Sigma-Aldrich) (selective) or SC -L (nonselective) growth media and allowed to grow to saturation. This process was repeated twice more for each of the two media. All three of the resulting saturated cultures for each selective and nonselective conditions were pooled and DNA was extracted to serve as template for cloning the TrpB fitness assay library, in addition to DNA extracted from generation 340 and 510 without downsampling. Note that this downsampling performed in yeast preceded the downsampling in *E. coli* necessary for proper sequencing coverage.

Pooled TrpB fitness assay of A20 mutants

A set of five strongly covarying mutations were selected based on data shown in Fig. 4D: A20V, A118V, F122S, I271V, and T292S. Additionally, the A20T mutation was selected as an alternative A20 mutation that did not covary strongly with this set of five mutations. All 48 (2^4 x 3^1) combinations of these six mutations were individually generated using overlap extension PCR to generate the full length TrpB variants followed by Golden Gate assembly to insert these variants into a genomic integration vector. The resulting constructs were genomically integrated into yeast strain OR-Y260. Ten individual colonies for each TrpB variant were first grown to saturation independently in nonselective media, then all 10 replicates for all variants were combined and passaged 1:100 into 100 ml of nonselective media. Upon saturation, this preculture was used to inoculate selective media to determine the relative enrichment of different TrpB variants. Specifically, the preculture was passaged 1:100 into SC -L with 1 mg/ml Trp and 400 μM indole, then grown to saturation. Genomic DNA harvested from the resulting saturated culture served as the postselection time point for sequencing and genomic DNA harvested from the initial culture used to inoculate selective media served as the preselection time point for sequencing. OD_{600} (the optical density of a sample at a wavelength of 600 nm) measurements were taken at the beginning of growth in selective media and at the end. The OD_{600} measurement at the beginning of growth was more specifically obtained by measuring the OD_{600} of the preculture and dividing that by 100 because the preculture was inoculated 1:100 into selective media.

We performed HTS on the pre- and postselection populations after the in vivo downsampled UMI PCR described above. Only sequencing reads that exactly matched one of the 48 manually constructed TrpBs (93% of reads) were included in read counts used for relative growth rate determination. From the UMI counts for each mutant TrpB and WT TrpB and with the pre- and postselection OD measurements, we were able to determine the relative growth rate difference for each mutant TrpB against WT TrpB as

$$g_i = \frac{r_i - r_{\mathrm{WT}}}{r_{\mathrm{WT}}}$$

where $r_i = \ln\left[\frac{N_i(t_1)}{N_i(t_0)}\right]$, $N_i(t_j) = \frac{\mathrm{OD}(t_j)\mathrm{UMI}_i(t_j)}{\sum_{i\in S}\mathrm{UMI}_i(t_j)}$, t_j signifies time point j, OD(t_j) is the measured OD for the entire population at time point j, $\mathrm{UMI}_i(t_j)$ is the UMI count for variant i at time point j, and S refers to the set of all 48 TrpBs.

To elaborate, we note that $N_i(t_j)$ is the population size for variant i at time point j in units OD. It is calculated by taking the UMI count for variant i at time point j as a fraction of the total UMI count for all variants at time point j, multiplied by the OD of the culture at time point j. We next note that the ratio $\left[\frac{N_i(t_1)}{N_i(t_0)}\right]$ represents the factor by which a given variant's population expanded from the start of growth in selective media at t_0 to the end at t_1. The natural log of this ratio, r_i, therefore corresponds to the (Malthusian parameter) × (elapsed time) for variant i. The relative growth rate difference between a mutant and WT organism can be defined as the difference between the Malthusian parameters for the mutant and the wild type divided by the Malthusian parameter for the wild type. This is equivalent to $\frac{r_i - r_{\mathrm{WT}}}{r_{\mathrm{WT}}}$, which we call g_i. This relative growth rate difference can be operationally treated as a selection coefficient. We show g_i for all 47 mutant TrpBs assayed in fig. S21.

Large scale pooled TrpB fitness assay

UMI-tagged evolved TrpB libraries and equivalent plasmids expressing two control TrpB sequences (*Tm*Triple and TrpB-003-1-A) were transformed into yeast strain OR-Y260 and plated onto -LH media, resulting in 40-fold coverage of the ~120-thousand-member library. Colonies were harvested and the library was spiked with each of the two control TrpB-expressing yeast strains at a 1:1000 control: library ratio. DNA was extracted from the resulting library to serve as the 0th time point. This library was also passaged 1:100 into 50 ml of either SC -L (nonselective), SC -LW + 400 μM indole (weakly selective), or SC -LW + 25 μM indole (strongly selective) and grown to saturation. This passaging and growth was repeated for each of the three growth conditions five times for a total of six passages. Of these, DNA was extracted from passages 1, 2, 5, and 6 to serve as additional time points. DNA from all time points were used as templates for PCR amplification of only the UMI and barcode regions for HTS.

Enrichment scores were calculated following the procedure described in Rubin *et al.* (77), albeit with two modifications to account for disparate sequencing coverage over multiple time points. First, counts of each UMI at each time point were normalized to the average count of all UMIs that persisted throughout all time points before log transformation and weighted linear regression. Second, regression weights were multiplied by this average count. All enrichment calculations were performed by the Maple pipeline within the script enrichment.py.

HTS analysis

High-accuracy consensus sequence generation, alignment, demultiplexing, genotype identification, mutation rate analysis, and basic dataset visualization were performed by version v0.10.4 of the Maple pipeline, which uses the Snakemake workflow management library, and is available on Zenodo (78). Parameters and settings for Maple analyses for each dataset and all other code used for analysis are also available on Zenodo (71).

TrpB fitness prediction with TranceptEVE

A multiple sequence alignment (MSA) of natural TrpB subunits was created using five iterations of jackhmmer (*79*) to query the UniRef100 database with a bitscore of 0.9. Columns with more than 20% gaps were ignored, and we used a theta parameter of 0.8 to downweight sequences with more than 80% sequence homology, as described in Hopf *et al.* (*80*). We trained four separate EVE models using this MSA and used them to calculate the log probabilities of the mutated sequences. These scores were ensembled with the Tranception log probabilities for each sequence by taking a weighted average of 60% Tranception score and 40% EVE score.

Null hypothesis sequence simulation and analysis

A dataset of TrpB variants that contained random mutations representative of biases due to the relevant mutation preferences and WT sequence, but that were not subject to selective forces, was generated through a simple simulation. First, the number of synonymous mutations within the open reading frame (ORF) of each evolved TrpB sequence was approximated as the number of nucleotide mutations minus the number of nonsynonymous mutations. For each real genotype, a corresponding simulated genotype was generated by starting from the WT TrpB sequence used in the evolution experiment and stochastically sampling nucleotide mutations with probabilities determined by the mutation rates of the same polymerase used for TrpB evolution (BadBoy2) until the same number of synonymous mutations as the evolved genotype was reached. Additional information, such as the time point from which the real sequence was identified and the count of the genotype, was also replicated for the corresponding simulated sequence. To account for some minor strand-dependent mutational biases, mutation rates and spectrum calculated for a sequence in the same position and orientation (relative to the LEU2 gene) as TrpB were used to generate the simulated sequences.

To validate our assumption that synonymous mutations were minimally subject to selective forces and could therefore be used as a neutral molecular clock in the simulation of mutation accumulation, we computationally generated mutagenized sequences using a bulk mutation process. In silico mutagenesis was carried out as above, but the number of synonymous mutations present in the evolved sequences was not used to determine the extent of mutagenesis. Instead, the projected average number of mutations per sequence was estimated using the overall nucleotide substitution rate of BadBoy2 and the estimated number of generations at each time point, and 10,000 sequences per time point were mutagenized to match this average of mutations per sequence. The resulting distribution of mutations per sequence in these sequences is shown in fig. S17 as "projected." All other figures referring to in silico mutagenized sequences refer to the simulation described in the previous paragraph.

Simulated and evolved sequences were analyzed identically. Isoelectric point (pI) and hydrophobicity index (gravy) were both calculated using the ProtParam module in BioPython (*81*). Note that the 6x histidine tag present on the TrpB sequence used as a starting point for evolution was included in the calculation of pI for all evolved TrpB sequences but not for the calculation of pI of the native TrpA-TrpB complex shown in Fig. 3E. Mesophilic adaptation mutations were selected from Haney *et al.* (*13*) as the inverse of the 17 mesophile to thermophile "Replacements most biased in number" with $P < 0.005$. Alternative sets of amino acid replacements were not evaluated.

Computational lineage downsampling

Lineage barcodes were identified using the demultiplexing feature within the Maple pipeline. The 100 most frequently observed lineage barcodes only from the first time point of TrpB evolution were used for all lineage analyses. Lineage barcodes for all remaining time points were assigned from this list of 100 barcodes, allowing for a nucleotide Hamming distance of 1. For global covariation analysis, all barcodes appearing in at least 100 sequences were identified, and 100 sequences from each lineage were randomly extracted for analysis of mutation frequencies. Covariation with specific mutations was performed similarly, except that only sequences that contained the specific mutation and were identified from specified time points were considered before randomly extracting sequences for analysis of mutation frequencies. To ensure random sampling did not bias results, this process was performed multiple times with virtually identical results.

REFERENCES AND NOTES

1. D. M. Blow, J. J. Birktoft, B. S. Hartley, Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **221**, 337–340 (1969). doi: 10.1038/221337a0; pmid: 5764436
2. G. Casari, C. Sander, A. Valencia, A method to predict functional residues in proteins. *Nat. Struct. Biol.* **2**, 171–178 (1995). doi: 10.1038/nsb0295-171; pmid: 7749921
3. J. A. Capra, M. Singh, Predicting functionally important residues from sequence conservation. *Bioinformatics* **23**, 1875–1882 (2007). doi: 10.1093/bioinformatics/btm270; pmid: 17519246
4. B. Reva, Y. Antipin, C. Sander, Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011). doi: 10.1093/nar/gkr407; pmid: 21727090
5. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009). doi: 10.1016/j.cell.2009.07.038; pmid: 19703402
6. J. W. McCormick, M. A. X. Russo, S. Thompson, A. Blevins, K. A. Reynolds, Structurally distributed surface sites tune allosteric regulation. *eLife* **10**, e68346 (2021). doi: 10.7554/eLife.68346; pmid: 34132193
7. F. Morcos *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011). doi: 10.1073/pnas.1111471108; pmid: 22106262
8. D. S. Marks *et al.*, Protein 3D structure computed from evolutionary sequence variation. *PLOS ONE* **6**, e28766 (2011). doi: 10.1371/journal.pone.0028766; pmid: 22163331
9. S. W. Lockless, R. Ranganathan, Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999). doi: 10.1126/science.286.5438.295; pmid: 10514373
10. E. Rivas, J. Clements, S. R. Eddy, A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45–48 (2017). doi: 10.1038/nmeth.4066; pmid: 27819659
11. I. N. Shindyalov, N. A. Kolchanov, C. Sander, Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Eng. Des. Sel.* **7**, 349–358 (1994). doi: 10.1093/protein/7.3.349; pmid: 8177884
12. D. Altschuh, A. M. Lesk, A. C. Bloomer, A. Klug, Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J. Mol. Biol.* **193**, 693–707 (1987). doi: 10.1016/0022-2836(87)90352-4; pmid: 3612789
13. P. J. Haney *et al.*, Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3578–3583 (1999). doi: 10.1073/pnas.96.7.3578; pmid: 10097079
14. G. Gianese, P. Argos, S. Pascarella, Structural adaptation of enzymes to low temperatures. *Protein Eng. Des. Sel.* **14**, 141–148 (2001). doi: 10.1093/protein/14.3.141; pmid: 11342709
15. I. N. Berezovsky, E. I. Shakhnovich, Physics and evolution of thermophilic adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12742–12747 (2005). doi: 10.1073/pnas.0503890102; pmid: 16120678
16. E. M. Marcotte, I. Xenarios, A. M. van Der Bliek, D. Eisenberg, Localizing proteins in the cell from their phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12115–12120 (2000). doi: 10.1073/pnas.220399497; pmid: 11035803
17. E. Shakhnovich, V. Abkevich, O. Ptitsyn, Conserved residues and the mechanism of protein folding. *Nature* **379**, 96–98 (1996). doi: 10.1038/379096a0; pmid: 8538750
18. R. V. Wolfenden, P. M. Cullis, C. C. F. Southgate, Water, protein folding, and the genetic code. *Science* **206**, 575–577 (1979). doi: 10.1126/science.493962; pmid: 493962
19. D. Collias, C. L. Beisel, CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.* **12**, 555 (2021). doi: 10.1038/s41467-020-20633-y; pmid: 33483498
20. J. Murciano-Calles, D. K. Romney, S. Brinkmann-Chen, A. R. Buller, F. H. Arnold, A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chem. Int. Ed.* **55**, 11577–11581 (2016). doi: 10.1002/anie.201606242; pmid: 27510733
21. F. Baier *et al.*, Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *eLife* **8**, e40789 (2019). doi: 10.7554/eLife.40789; pmid: 30719972
22. G. Gasiunas *et al.*, A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nat. Commun.* **11**, 5512 (2020). doi: 10.1038/s41467-020-19344-1; pmid: 33139742
23. M. H. Medema, T. de Rond, B. S. Moore, Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* **22**, 553–571 (2021). doi: 10.1038/s41576-021-00363-7; pmid: 34083778
24. D. L. Trudeau, M. A. Smith, F. H. Arnold, Innovation by homologous recombination. *Curr. Opin. Chem. Biol.* **17**, 902–909 (2013). doi: 10.1016/j.cbpa.2013.10.007; pmid: 24182747
25. A. Crameri, S. A. Raillard, E. Bermudez, W. P. C. Stemmer, DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998). doi: 10.1038/34663; pmid: 9440693
26. A. J. Riesselman, J. B. Ingraham, D. S. Marks, Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **15**, 816–822 (2018). doi: 10.1038/s41592-018-0138-4; pmid: 30250057
27. J. Frazer *et al.*, Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021). doi: 10.1038/s41586-021-04043-8; pmid: 34707284
28. A. Rives *et al.*, Biological structure and function emerge from scaling unsupervised learning to 250 million protein

sequences. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016239118 (2021). doi: 10.1073/pnas.2016239118; pmid: 33876751
29. J. E. Shin *et al*., Protein design and variant prediction using autoregressive generative models. *Nat. Commun.* **12**, 2403 (2021). doi: 10.1038/s41467-021-22732-w; pmid: 33893299
30. D. H. Bryant *et al*., Deep diversification of an AAV capsid proteins by machine learning. *Nat. Biotechnol.* **39**, 691–696 (2021). doi: 10.1038/s41587-020-00793-4; pmid: 33574611
31. A. Madani *et al*., Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**, 1099–1106 (2023). doi: 10.1038/s41587-022-01618-2; pmid: 36702895
32. T. A. Hopf *et al*., Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **149**, 1607–1621 (2012). doi: 10.1016/j.cell.2012.04.012; pmid: 22579045
33. J. Jumper *et al*., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
34. K. Tunyasuvunakool *et al*., Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021). doi: 10.1038/s41586-021-03828-1; pmid: 34293799
35. W. Makałowski, M. S. Boguski, Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9407–9412 (1998). doi: 10.1073/pnas.95.16.9407; pmid: 9689093
36. M. Nei, P. Xu, G. Glazko, Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2497–2502 (2001). doi: 10.1073/pnas.051611498; pmid: 11226267
37. M. A. Stiffler *et al*., Protein Structure from Experimental Evolution. *Cell Syst.* **10**, 15–24.e5 (2020). doi: 10.1016/j.cels.2019.11.008; pmid: 31838147
38. G. Rix *et al*., Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **11**, 5644 (2020). doi: 10.1038/s41467-020-19539-6; pmid: 33159067
39. M. S. Morrison, C. J. Podracky, D. R. Liu, The developing toolkit of continuous directed evolution. *Nat. Chem. Biol.* **16**, 610–619 (2020). doi: 10.1038/s41589-020-0532-y; pmid: 32444838
40. R. S. Molina *et al*., In vivo hypermutation and continuous evolution. *Nat. Rev. Methods Primers* **2**, 36 (2022). doi: 10.1038/s43586-022-00119-5; pmid: 37073402
41. A. Ravikumar, A. Arrieta, C. C. Liu, An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014). doi: 10.1038/nchembio.1439; pmid: 24487693
42. A. Ravikumar, G. A. Arzumanyan, M. K. A. Obadi, A. A. Javanpour, C. C. Liu, Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **175**, 1946–1957.e13 (2018). doi: 10.1016/j.cell.2018.10.021; pmid: 30415839
43. Z. Zhong, A. Ravikumar, C. C. Liu, Tunable Expression Systems for Orthogonal DNA Replication. *ACS Synth. Biol.* **7**, 2930–2934 (2018). doi: 10.1021/acssynbio.8b00400; pmid: 30408954
44. J. D. García-García *et al*., Using continuous directed evolution to improve enzymes for plant applications. *Plant Physiol.* **188**, 971–983 (2022). doi: 10.1093/plphys/kiab500; pmid: 34718794
45. E. D. Jensen *et al*., Integrating continuous hypermutation with high-throughput screening for optimization of *cis*,*cis*-muconic acid production in yeast. *Microb. Biotechnol.* **14**, 2617–2626 (2021). doi: 10.1111/1751-7915.13774; pmid: 33645919
46. A. A. Javanpour, C. C. Liu, Evolving Small-Molecule Biosensors with Improved Performance and Reprogrammed Ligand Preference Using OrthoRep. *ACS Synth. Biol.* **10**, 2705–2714 (2021). doi: 10.1021/acssynbio.1c00316; pmid: 34597502
47. A. Wellner *et al*., Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat. Chem. Biol.* **17**, 1057–1064 (2021). doi: 10.1038/s41589-021-00832-4; pmid: 34168368
48. E. P. Harvey *et al*., An in silico method to assess antibody fragment polyreactivity. *Nat. Commun.* **13**, 7554 (2022). doi: 10.1038/s41467-022-35276-4; pmid: 36477674
49. E. Vallina Estrada, N. Zhang, H. Wennerström, J. Danielsson, M. Oliveberg, Diffusive intracellular interactions: On the role of protein net charge and functional adaptation. *Curr. Opin. Struct. Biol.* **81**, 102625 (2023). doi: 10.1016/j.sbi.2023.102625; pmid: 37331204
50. S. E. Luria, M. Delbrück, Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491–511 (1943). doi: 10.1093/genetics/28.6.491; pmid: 17247100
51. P. L. Foster, Methods for determining spontaneous mutation rates. *Methods Enzymol.* **409**, 195–213 (2006). doi: 10.1016/S0076-6879(05)09012-9; pmid: 16793403
52. M. Jain, H. E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016). doi: 10.1186/s13059-016-1103-0
53. P. J. Zurek, P. Knyphausen, K. Neufeld, A. Pushpanath, F. Hollfelder, UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Nat. Commun.* **11**, 6023 (2020). doi: 10.1038/s41467-020-19687-9; pmid: 33243970
54. S. M. Karst *et al*., High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods* **18**, 165–169 (2021). doi: 10.1038/s41592-020-01041-y; pmid: 33432244
55. H. A. Orr, The Rate of Adaptation in Asexuals. *Genetics* **155**, 961–968 (2000). doi: 10.1093/genetics/155.2.961
56. P. J. Gerrish, A. Colato, P. D. Sniegowski, Genomic mutation rates that neutralize adaptive evolution and natural selection. *J. R. Soc. Interface* **10**, 20130329 (2013). doi: 10.1098/rsif.2013.0329; pmid: 23720539
57. M. F. Dunn, Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. *Arch. Biochem. Biophys.* **519**, 154–166 (2012). doi: 10.1016/j.abb.2012.01.016; pmid: 22310642
58. A. R. Buller *et al*., Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14599–14604 (2015). doi: 10.1073/pnas.1516401112; pmid: 26553994
59. D. S. Goodsell, L. Autin, A. J. Olson, Illustrate: Software for Biomolecular Illustration. *Structure* **27**, 1716–1720.e1 (2019). doi: 10.1016/j.str.2019.08.011; pmid: 31519398
60. A. R. Buller *et al*., Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* **140**, 7256–7266 (2018). doi: 10.1021/jacs.8b03490; pmid: 29712420
61. M. A. Maria-Solano, J. Iglesias-Fernández, S. Osuna, Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *J. Am. Chem. Soc.* **141**, 13049–13056 (2019). doi: 10.1021/jacs.9b03646; pmid: 31356074
62. R. Orij, J. Postmus, A. Ter Beek, S. Brul, G. J. Smits, *In vivo* measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in *Saccharomyces cerevisiae* reveals a relation between intracellular pH and growth. *Microbiology* **155**, 268–278 (2009). doi: 10.1099/mic.0.022038-0; pmid: 19118367
63. H. Wennerström, E. Vallina Estrada, J. Danielsson, M. Oliveberg, Colloidal stability of the living cell. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10113–10121 (2020). doi: 10.1073/pnas.1914599117; pmid: 32284426
64. L. Xiang, R. Yan, K. Chen, W. Li, K. Xu, Single-Molecule Displacement Mapping Unveils Sign-Asymmetric Protein Charge Effects on Intraorganellar Diffusion. *Nano Lett.* **23**, 1711–1716 (2023). doi: 10.1021/acs.nanolett.2c04379; pmid: 36802676
65. M. Leander, Y. Yuan, A. Meger, Q. Cui, S. Raman, Functional plasticity and evolutionary adaptation of allosteric regulation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25445–25454 (2020). doi: 10.1073/pnas.2002613117; pmid: 32999067
66. A. Shaw *et al*., Removing bias in sequence models of protein fitness. bioRxiv 2023.09.28.560044 [Preprint] (2023); https://doi.org/10.1101/2023.09.28.560044.
67. P. Notin *et al*., TranceptEVE: Combining Family-specific and Family-agnostic Models of Protein Sequences for Improved Fitness Prediction. bioRxiv 2022.12.07.519495 [Preprint] (2022); https://doi.org/10.1101/2022.12.07.519495.
68. B. L. Hie *et al*., Efficient evolution of human antibodies from general protein language models. *Nat. Biotechnol.* **42**, 275–283 (2024). doi: 10.1038/s41587-023-01763-2; pmid: 37095349
69. Z. Zhong *et al*., Automated Continuous Evolution of Proteins *in Vivo*. *ACS Synth. Biol.* **9**, 1270–1276 (2020). doi: 10.1021/acssynbio.0c00135; pmid: 32374988
70. E. A. DeBenedictis *et al*., Systematic molecular evolution enables robust biomolecule discovery. *Nat. Methods* **19**, 55–64 (2022). doi: 10.1038/s41592-021-01348-4; pmid: 34969982
71. Iiusynevolab, Iiusynevolab/OrthoRep_Rix_2024: v1, version v1, Zenodo (2024); https://doi.org/10.5281/zenodo.11187038.
72. E. Alani, L. Cao, N. Kleckner, A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. *Genetics* **116**, 541–545 (1987). doi: 10.1093/genetics/116.4.541; pmid: 3305158
73. O. W. Ryan, J. H. D. Cate, in *Methods in Enzymology*, vol. 546, J. A. Doudna, E. J. Sontheimer, Eds. (Elsevier, ed. 1, 2014), pp. 473–489.
74. R. Volden *et al*., Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9726–9731 (2018). doi: 10.1073/pnas.1806447115; pmid: 30201725
75. R. T. Oliynyk, G. M. Church, Efficient modification and preparation of circular DNA for expression in cell culture. *Commun. Biol.* **5**, 1393 (2022). doi: 10.1038/s42003-022-04363-z; pmid: 36543890
76. Y. Zhang, N. A. Tanner, Isothermal Amplification of Long, Discrete DNA Fragments Facilitated by Single-Stranded Binding Protein. *Sci. Rep.* **7**, 8497 (2017). doi: 10.1038/s41598-017-09063-x; pmid: 28819114
77. A. F. Rubin *et al*., A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017). doi: 10.1186/s13059-017-1272-5; pmid: 28784151
78. G. Rix, gordonrix/maple: Rix_et_al_2024, version v0.10.4, Zenodo (2024); https://doi.org/10.5281/zenodo.11179609.
79. L. S. Johnson *et al*., Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010). doi: 10.1186/1471-2105-11-431; pmid: 20718988
80. T. A. Hopf *et al*., Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017). doi: 10.1038/nbt.3769; pmid: 28092658
81. P. J. A. Cock *et al*., Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009). doi: 10.1093/bioinformatics/btp163; pmid: 19304878

ACKNOWLEDGMENTS

We thank members of the Liu group for materials and thoughtful discussions. **Funding:** This study was supported by NIH NIGMS R35GM136297 (C.C.L.), NIH NCI R01CA260415 (C.C.L. and D.S.M.), and a Hewitt Foundation for Medical Research Postdoctoral Fellowship (R.L.W.). **Author contributions:** Conceptualization: G.R. and C.C.L. Methodology: G.R., C.C.L., V.J.H., and A.S. Investigation: G.R., R.L.W., V.J.H., A.S., and A.P. Visualization: G.R. and V.J.H. Funding acquisition: C.C.L. and D.S.M. Supervision: C.C.L. and D.S.M. Writing – original draft: G.R. and C.C.L. Writing – review & editing: G.R., C.C.L., R.L.W., V.J.H., A.S., and D.S.M. **Competing interests:** A provisional patent on this work has been filed with C.C.L., G.R., and R.L.W. as inventors. C.C.L. is a cofounder of K2 Biotechnologies, Inc., which uses OrthoRep for protein engineering. D.S.M. is an advisor for Dyno Therapeutics, Octant, Jura Bio, Tectonic Therapeutic, and Genentech and is a cofounder of Seismic Therapeutic. The authors declare no other competing interests. **Data and materials availability:** All data are provided or available online, and materials generated for this study are available upon request to the corresponding author. All code and scripts are available on Zenodo, including the Maple sequencing analysis pipeline developed for this study (*78*), additional sequencing analysis scripts, experiment-specific Maple parameters, and analyzed HTS data (*71*). Raw HTS data are available on the NCBI SRA with accession no. PRJNA1050257.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adm9073
Supplementary Text
Figs. S1 to S21
Tables S1 to S8
References (*82*–*86*)
MDAR Reproducibility Checklist

Submitted 12 November 2023; accepted 10 September 2024
10.1126/science.adm9073